USING WIFI MOBILITY DATA FOR MODELING COVID-19 ON UNIVERSITY CAMPUSES

A Dissertation
Presented to
The Academic Faculty

By

Jiajia Xie

In Partial Fulfillment
of the Requirements for the Degree of
Master of Science in
Computational Science and Engineering
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology

May 2021

USING WIFI MOBILITY DATA FOR MODELING COVID-19 ON UNIVERSITY CAMPUSES

Thesis committee:

Dr. B. Aditya Prakash
School of Computational Science and Engineering, College of Computing
Georgia Institute of Technology

Dr. Lauren N. Steimle
H. Milton Stewart School of Industrial and Systems Engineering, College of Engineering
Georgia Institute of Technology

Dr. Pinar Keskinocak
H. Milton Stewart School of Industrial and Systems Engineering, College of Engineering
Georgia Institute of Technology

Date approved: April 10, 2021

For my parents

ACKNOWLEDGMENTS

First, I would like to express the most profound appreciation to my advisors Prof. B. Aditya Prakash and Prof. Lauren N. Steimle, who have been my nice friends and mentors. I really enjoyed being a student of them as we explored exciting research questions about Covid-19 under the global pandemic. Their generosity of time, patient, and encouragement eventually help me in preparing this work.

Also, I want to thank my thesis committee member Prof. Pinar Keskinocak for her timely and valuable feedback and suggestions on this thesis.

In addition, I would like to thank Prof. Munmun De Choudhury, especially for the opportunity of working in Socweb, which is literally the starting point of my research career. I would also like to thank Prof. Gregory Abowd for being an excellent organizer and collaborator from Ubicomp.

Special thanks are due to the friends and colleagues who made this work possible. Vedant Das Swain, Mehrab Bin Morshed, and Alex Rodriguez were invaluable both as friends and as sounding boards for some of my more outlandish ideas.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

EN Course enrollment data

NABM network agent-based Models

WIMOB WiFi Mobility

LC_{BCen} Localized closure policies with Betweenness Centrality

LC_{ECen} Localized closure policies with Eigenvector Centrality

LC_{LCen} Localized closure policies with Load Centrality

LC_{PRank} Localized closure policies with Pagerank

LC Localized closure policy

RI remote instruction policies

S1 Persistent Scenario

S2 Non-residential Avoidance

S3 Complete Avoidance

Berkeley University of California, Berkeley

GT Georgia Institute of Technology

UIUC University of Illinois at Urbana-Champaign

SUMMARY

Infectious diseases, like mumps, flu, or measles, can cause devastating impacts on universities. To protect the community's health, schools must learn how to operate during epidemics. In the light of Covid-19, for instance, the universities in the U.S. have struggled to bring students, staff, and faculties back to campuses. On the one hand, these schools are often hotspots for outbreaks. On the other hand, a long-term local lockdown will likely incur losses of financial income for the school and the local businesses due to diminishing enrollment and limited visits to campuses and their surrounding neighborhoods

To meet the reopening goal, the schools must assess the ongoing epidemic of Covid-19 on campus and design operational plans with more robust and accurate information than the data provided by other local agencies as support. Frequently asked questions from the perspective of campus officials are: How can we predict potential outcomes of disease spread? How can we evaluate strategies to control the epidemic? Which groups of individuals and locations are particularly vulnerable to Covid-19? How can we prioritize the testing program among individuals active on campus? Answering those questions typically involves disease modeling since models help us abstract the disease dynamics and reason more about the mechanism of disease transmissions among the community.

This thesis targets several natural and fundamental problems for universities during the Covid-19 pandemic using human mobility data. We propose using the on-campus WiFi infrastructure to understand human mobility and approximate contact networks among individuals on campus. When an individual accesses the WiFi on campus, their device sends a request to a WiFi access point which creates a record that the device was connected to the WiFi network. From these logs, we can determine when and for how long that individual was connected to the WiFi through a particular access point and infer the location of that individual to the level of a room on campus. More formally, the logs give us a bipartite network between users and WiFi access points across different time-stamps, defined

as WiFi Mobility data. Each connection of a user to a log at any time will be recorded as an edge in the bipartite network. Using a projection of this bipartite network, we can infer which individuals come into close proximity of each other on campus. We construct and validate a network-based simulation model of Covid-19 on university campuses using Wifi mobility data to approximate the contact network among individuals. Then, we design and evaluate two novel methods for improving decision-making powered by the WiFi mobility data with the model constructed. The first method outputs a more granular and localized closure policy, causing more effective disease intervention outcomes but less burdensome to individuals and schools. The second can discover likely chains of transmission among individuals and missing infections on campus given the current testing report data.

CHAPTER 1
INTRODUCTION

1.1 Motivation and Overview

Outbreaks of infectious diseases, like mumps, influenza, or measles, can cause devastating impacts on universities [1, 2]. The Covid-19 pandemic has exposed many challenges that universities face to bring students, staff, and faculties back to campuses during an infectious disease outbreak or epidemic. On the one hand, colleges may opt fear the health consequences associated with outbreaks which can negatively impact the health of the students, faculty, staff, and broader community [3]. For example, there are reports that the number of infections significantly increases after students return to campuses [4]. Outbreaks on college campuses can also contribute to overwhelming the regional medical system [5]. On the other hand, there are reasons that colleges may wish to avoid completely shutting down during an outbreak. Universities that undergo long-term closures will likely incur losses of financial income for both the schools due to diminished enrollment and the local economy due to limited visits to campus and their surrounding neighborhoods [6, 7]. Moreover, students continuously taking classes online under university closures will likely experience more negative emotion such as stress, anxiety, and nervousness, which will further affect their learning outcomes [8, 9].

For the above reasons, colleges have strong incentives to find ways to safely operate during outbreaks and epidemics. For instance, when the Covid-19 emerged, colleges sought to design or improve operational plans to minimize the size of potential outbreak. Frequently asked questions from the perspective of campus administrators are: How can we predict potential outcomes of disease spread? How can we evaluate strategies to control the epidemic? Which groups of individuals and locations are particularly vulnerable to infec-

tion? How should we prioritize individuals for testing on campus? How can we discover potential chains of transmission and missing infections?

Disease modeling can be helpful for answering these questions. Disease modeling helps abstract the disease dynamics and reason about the mechanism of disease transmission among a community. Recent mathematical models of Covid-19 in the literature have evaluated non-pharmaceutical intervention strategies on university campuses under various scenarios of testing strategies or students behaviors in response to social distancing [10, 11, 12]. However, these models are usually parameterized using summary statistics from data on the campus community, such as statistics on enrollment. Further, these models are typically limited by their simple assumptions about human behavior, such as a well-mixing population (or a clique network [13]). Thus, the experimental results based on these models may be inaccurate. On the other hand, Riti and Edward propose using network-based models, a type of computer agent-based simulation, network agent-based Models (NABM), to avoid these assumptions [14, 15]. These methods typically require an approximated contact network that represents which agents interact in terms of real-world data [16]. For example, Riti and Nicole et al. draw networks describing the connections between building blocks within small residential colleges based on campus maps; however these blocks are also estimated from summary data [14]. In addition, multiple studies highlight the potential of course and residence enrollment data that can be reshaped into contact networks among individuals [17, 18, 19]. However, these networks are static in general, with a limited scope of information given from registration data updated non-frequently. Due to the nature of the data used to construct these existing models, the closure policy interventions considered have been limited to broad shutdowns and class closures which can be burdensome to individuals and schools.

This thesis targets several natural and fundamental problems for universities related to infectious disease control motivated by the Covid-19 pandemic. We demonstrate that using human mobility data harnessed from the on-campus managed WiFi network is useful

for answering these questions. The data enables us to model the disease dynamic realistically on campus and evaluate data mining algorithms for answering related epidemiological questions. The study will draw upon and contribute to knowledge from multiple fields related to the interdisciplinary nature of the issues, including computer science, public health, and operation research.

1.2 Thesis statement

WiFi mobility data can assist universities in designing and evaluating strategies for controlling infections on campus.

We propose using the on-campus WiFi infrastructure to understand human mobility and approximate contact networks among individuals on campus. When an individual accesses the WiFi on campus, their device sends a request to a WiFi access point which creates a record that the device was connected to the WiFi network. From these logs, we can determine when and for how long that individual was connected to the WiFi through a particular access point and infer the location of that individual to the level of a room on campus. More formally, the logs give us a bipartite network between users and WiFi access points across different time-stamps, defined as, WiFi Mobility (WIMOB). Each connection of an user to a log at any time will be recorded as an edge in the bipartite network. Using a projection of this bipartite network, we can infer which individuals come into close proximity of each other on campus.

Using WIMOB to approximate the contact network among individuals, we construct and validate a network-based simulation model of Covid-19 on a university campus. Then, we design and evaluate two novel methods for improving decision-making powered by the WiFi mobility data with the model constructed. The first method outputs a more granular and localized closure policy, which is more effective in terms of disease intervention outcomes but also less burdensome to individuals and schools. The second method can reliably reconstruct likely chains of transmissions over time, given the current noisy testing

data.

We organize the thesis in the following outline. In Chapter 1, we present the overview and introduction for the thesis. In Chapter 2, we define, calibrate, and validate our NABM model. In Chapter 3, we design and evaluate our closure policies. In Chapter 4, we demonstrate that WIMOB can be used to discover the past chains of transmission among individuals and missing infections. In Chapter, we present the conclusion and discuss limitations and future works.

1.3 Chapter 2

In this chapter, we develop an NABM for modeling Covid-19 on university campuses, with the projected collocation networks from on-campus WiFi mobility data WIMOB as the contact network. NABM is a simulation-based model that can describe the disease dynamics realistically with no assumption about the interactions pattern among population. Moreover, NABM leverages a commonly-used SEIR framework to describe epidemiological properties of Covid-19, such as asymptomatic infections and incubation period. The model also reflects some commonly used operation guideline of universities for symptomatic cases such as quarantine in places on campus. We calibrate NABM to different surveillance reports of universities. The validation results demonstrate the model for simulating the disease dynamics on-campus and accurately predicts the future cases. Given the calibrated NABM, We will further use it to experiment with different policies and approaches in the following sections.

1.4 Chapter 3

This chapter is devoted to introducing a more effective and less burdensome policy, Localized closure policy (LC), based on WIMOB for a university to control the spread of Covid-19 on campus. In general, universities typically adapt remote instruction policies (RI), such as removing classes into online mode based on course enrollment to reduce

potential contact, which has side effects on campus by hampering the local economy, students' learning outcomes, and community well-being. We will demonstrate that university policymakers can mitigate these tradeoffs by leveraging WIMOB to learn community mobility and explore more granular policies like LC. WIMOB can construct contact networks that capture behavior in various spaces, highlighting new potential transmission pathways and temporal variation in contact behavior. Additionally, WIMOB enables us to design LC policies that close super-spreader locations at the granular level of rooms, suites, or dooms on campus. On simulating disease spread with contact networks from WIMOB, we find that LC maintains the same reduction in cumulative infections as RI while showing a greater reduction in peak infections and internal transmission. Moreover, LC reduces campus burden by closing fewer locations, forcing fewer students into completely online schedules, and requiring no additional isolation. WIMOB can empower universities to conceive and assess various closure policies to prevent future outbreaks.

1.5 Chapter 4

This chapter is devoted to showing that university officials can utilize WIMOB to identify likely sources and chains of transmissions causing an outbreak of infectious diseases on campus. Universities might want to know who the outbreak sources are and how the disease transmits when the disease spread out on campus for some time. In particular, they are interested in reconstructing the epidemic by discovering the sources and flows causing the outbreak among individuals on university campuses. By doing so, universities can identify groups of individuals and locations at risk, they can prioritize testing for people, or they can discover people who have been likely infected but not captured by testing. We present a problem formulation with an existing efficient algorithm for solving the problem, and the inputs are similar to the real-world data available to us: (1) a time-varying collocation network and (2) a noisy surveillance data reporting the number of positive cases for residential groups. Moreover, we validate the entire framework by instances from the

calibrated NABM we have designed in previous chapters. Finally, we present a case study of Covid-19 in one of the university campuses, Georgia Institute of Technology (GT), in Fall 2020 semester.

1.6 Chapter 5

The thesis concludes with Chapter 5 which presents a summary of the most important findings, discussion, and potential directions for future research that stems from this work.

CHAPTER 2

COVID-19 MODELING

2.1 Overview

Mechanistic epidemiology models are valuable tools for predicting the likely outcome of disease spreading among the targeted population and assisting health policy development for intervention and control [20]. However, modeling infectious diseases, like Covid-19, on university campuses remains challenging concerning the biological properties of the virus are imperfectly understood in an early stage, and the campus community has a different behavioral pattern in general. One example is mathematical models, where researchers typically make many assumptions [21]. These models are constrained by what we know and what we assume [13]. For instance, they might apply a hypothetical constant rate of contact among a well-mixing individuals (anyone can have contacts with anyone else) in the systems of equations [10, 11, 12]. These assumptions they made are generally far away from the reality concerning a much more sparse population network varying corresponding to different stages of a semester, see the right figure of Figure 2.2.

With the rise of computer power, agent-based simulation, has been adopted in epidemiology modeling [16]. The combination of agent-based simulation and network science applying in epidemiology can yield a more effective and flexible type of models, NABM, [22]. One feature of NABM is the model propagates the disease through physical contacts represented by edges from the network. Especially since NABM can capture the nonlinear relation between contacts among individuals and transmissions of the disease, NABM is a more natural description close to the real-world disease dynamics [23]. Note that NABM generally has no assumption over the contact patterns among the population. Instead, it requires a given contact network approximated from real-world data.

Universities can leverage much information to establish contact networks, such as course registration data or residential records [19, 17]. This chapter proposed using the on-campus WiFi data, WIMOB, to approximate the dynamic collocation network for constructing our NABM.

2.2 Contribution

In this chapter, we develop an network agent-based model, NABM for modeling Covid-19 on university campuses, with the projected collocation networks from on-campus WiFi mobility data WIMOB as the contact network. NABM is a simulation-based model that can describe the disease dynamics realistically with no assumption about the interactions pattern among population. Moreover, NABM leverages a commonly used SEIR framework to describe epidemiological properties of Covid-19, such as asymptomatic infections and incubation period. The model also reflects some commonly used operation guideline of universities for symptomatic cases such as quarantine in places on campus. We calibrate NABM with the contact network to different surveillance reports of universities. The validation results demonstrate a nice accuracy of the model for simulating the disease dynamics on-campus and predicting the future. Given the calibrated NABM, We will further use it to experiment with different policies and approaches in the following sections.

2.3 Epidemiological Modeling Concepts and SEIR Framework

Epidemiological models are valuable tools for human beings to analyze infectious diseases' mechanisms and to predict the outcome of spreading. In the 18th century, Daniel Bernoulli created the first mathematical model for studying the effect of variolation for a targeted population with smallpox [24]. Many current epidemiological models preserve the concept of both susceptible populations and the life-time immunity for infected survivors based on Bernoulli's observation from the smallpox population. It is common to see that epidemiological modeling always starts with empirical research about the disease dynamic and its

pathology [25]. For Covid-19, earlier clinical knowledge highlights that (1) there is a rare possibility for reinfection, (2) there is likely an incubation period for each patient, and (3) there is a possibility for patients contracted the disease without symptoms [26, 27]. Based on these observations, many existing compartmental frameworks can capture these observations in the model [28]. One example is the SEIR compartmental model based on differential equations that divides the entire population into four states: *susceptible*, *exposed*, *infected*, and *recovered* [29]. In particular, the *exposed* state [30]. We employ a fundamental SEIR model to demonstrate important concepts of this framework used for modeling Covid-19 in our study. In general, the SEIR model consists of the following four differential equations:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta IS \\
\frac{dE}{dt} &= \beta IS - \sigma E \\
\frac{dI}{dt} &= \sigma E - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned}$$

β is the rate of transmission for each contact between *susceptible* and *infected* population. σ is the rate of getting *infected* for every *exposed*. γ is the recovery rate of the *infected* population. The model has an initial condition such that $S + E + I + R = 1$. The reproductive number, R_0, of this model is defined as the ratio of:

$$R_0 = \frac{\beta}{\gamma}$$

If $R_0 > 1$, there will be an outbreak caused by the disease since the initial $\frac{dI}{dt} > 0$, otherwise the disease will die out. R_0 can also be interpreted as the number of infections caused by an extra infection within a population assumed everybody is *susceptible*. Note

that the differential-equation-based model requires three underline assumptions (1) Perfect mixing: meaning that every *infected* individual can infect every *susceptible*. (2) A constant population: the entire population has no new death and birth, and (3) the disease dynamics is a deterministic system. One can easily see that these assumptions are far from reality if we use the differential-equation-based SEIR to model the disease dynamics on university campuses. Nevertheless, recent studies about Covid-19 modeling commonly adopt the SEIR compartmental concept as the baseline framework for constructing various models according to their research purposes [31, 32, 33].

2.4 Data

The data we will leverage for modeling Covid-19 are defined as follows:

WiFi Mobility

We use data provided by the IT management facility at GT which accumulates WiFi access point (AP) logs over time. The primary use of WiFi network logs is for maintenance and security purposes. We mine these logs post-hoc to describe the mobility of individuals on campus, which we refer to as WIMOB. Here mobility is expressed by visits to certain locations that are demarcated by a corresponding AP. WIMOB can also describe dwelling (duration of visits) and collocation (dwelling in the presence of others around the same AP). Since we do not have access to other university's WiFi data, we use GT's WIMOB as the hypothetical contact network for these campuses and assume that the contact patterns shaped by WiFi network are similar to each other across different campuses.

The campus WiFi network spans 6959 APs distributed between 240 buildings (and some outdoor locations). We label APs according to which building they are inside, along with the closest room or space (e.g, hallway, lobby, suite, cafe, etc.). The AP may or may not reside inside the room, however, in most cases, only a single AP is associated with space. For less than 5% of the APs, the AP shared association to space with another AP. This

many-to-one mapping is typically in the case of large halls and auditoriums. We resolve such many-to-one associations by using APs as a proxy of the space they are associated with. Therefore, individuals connected to different APs in the same space will still be identified as collocated. Similarly, an individual could connect to the network with multiple devices. However, less than 1% logs show that a user is connected to multiple APs around the same time. Therefore, WIMOB is agnostic to which device connects to the APs to proxy the presence of the individual. For this study, we obtain the WiFi network logs retrospectively for all of Fall 2019, and the data for Fall 2020 was provided on a per-day basis. Each day, approximately $33,000$ different people connect their devices to the WiFi network on campus. Overall in Fall 2019, approximately $40,000$ different people connected to the campus network.

Note, that GT's managed WiFi network is not equipped with any Real-Time Location System (RTLS) [34, 35]. RTLS systems use Received Signal StrenGTh Indicator (RSSI) values from multiple neighboring APs to provide high precise localization of individuals in terms of time and space. However, deploying such systems requires surveying the entire network. Additionally, precision localization raises more privacy concerns. These factors together make it challenging for universities to justify the deployment of RTLS, unlike small retail settings that can monetize RTLS insights directly (e.g., insights on footfall can be tied to improving revenue).

Inferring location from Logs

WIMOB is our approach to describe contact between people and movement of people between locations. The first step requires using WiFi network logs to infer when individuals were at specific locations on campus by determining when devices were connected to the corresponding APs. Our system mines the WiFi network logs that are populated via the Simple Network Management Protocol (SNMP) — a standard and widely used monitoring protocol to organize device association behavior to a WiFi network. Periodic SNMP up-

dates can be caused either by poll requests to the APs that log which devices are associated with it at that time. However, devices can appear invisible to detached from an AP for multiple reasons, for example, when devices are idle. Otherwise. SNMP updates can occur whenever a new device connects, which is typical when individuals move between APs. Our approach exploits this factor to first mine periods when individuals are moving, then identify periods of dwelling between movements, and finally determine collocation when two or more individuals are dwelling near the same AP. This system follows from other studies that mine WiFi logs [36, 37] and the detailed processing pipeline and evaluation is presented in [38]. This system to infer collocations has been tested against lecture attendance and reports a high precision of 0.89, but a relatively lower specificity of 0.79 [38]. While it is not likely to show false-positives, it has a possibility to erroneously mark people absent from a location even though they were there. However, for the purposes of our study, a contact network is made over an entire day and it only needs a single collocation instance for us to consider contact. And therefore we believe this limitation would not significantly affect our models.

Characterizing Logs as Contact and Movement Networks

After inferring where an individual is located on campus, we represent the entire community behavior as graphs. We describe a bipartite graph, K, that shows when a user is at a given location on campus (Figure 2.1). This bipartite graph has edges connecting a set of m people, P, to a set of n locations, L. An individual can have multiple edges connecting to the location if they visited that location multiple times (e.g., t_1, t_2). The edge data contains the start and end times of these dwelling periods. For these bipartite graphs, we make a projection on set P to describe collocation. This projection graph, G, contains an edge between users if they were visiting the same location during overlapping times. Since we do not use RTLS, our approach can only identify if people were in the vicinity of the same AP, but does not describe the distance between them. However, it can reasonably de-

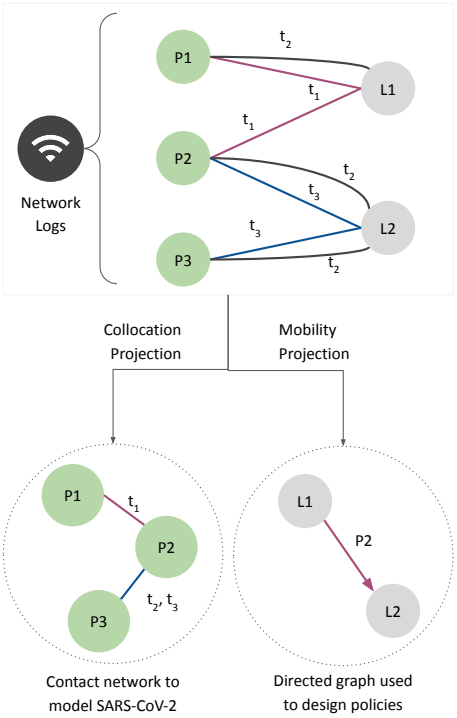


Figure 2.1: In a managed network, SNMP updates the logs by describing device association to an AP at a certain timestamp. WIMOB mines these logs to characterize mobility as a bipartite graph. The nodes are partitioned to describe people nodes (e.g., $P1$, $P2$) connected to locations nodes (e.g., $L1$, $L2$). Every edge across the partition describes people visiting locations on campus during different times (e.g., t_1, t_2). Projecting the bipartite on people nodes helps construct a contact network (e.g., $P1$ and $P2$ were collocated at $L1$ at t_1), while projecting it on locations helps construct a directed movement graph ($P2$ dwelled at $L1$ and then at $L2$).

termine collocation in the same room [38]. Since our study is limited to localizing people indoors, we adapt the definition of *proximate contact* [39] where people might be "more than 6 feet but in the same room for an extended period". In our work, we use a lower bound threshold of 40 minutes to determine proximate contact. Therefore, individuals are only considered in contact when they are collocated in a room for 40 minutes or more. This threshold was set up to account for typical lecture duration on campus (for standard 3-credit hour courses taught 3 times a week). Every edge between two individuals contains a list of locations where they were possibly in contact. G forms the basis of the contact-network that we use an agent-based model to simulate. Alternatively, we also make a projection on the set L. This projection is a directed graph, H, where an edge from L_i to L_j represents

movement from the first location to the next within a span of 60 minutes. GT's large urban campus with pedestrian pathways and motorized transit services enables direct movement between any two places on campus within the threshold. The 60 minutes threshold helps discount erroneously labeling returning from outside campus (e.g., non-residential students visiting two different locations between 2 days). H effectively describes how locations are connected and which locations could be more conducive to attracting and disseminating the virus. As a consequence, the H helps inform policy design. We compute the bipartite graph and its projections for each day of the semester.

Asymptomatic surveillance testing data

We calibrated the NABM using the publicly reported positivity rate on different campuses, (GT, University of Illinois at Urbana-Champaign (UIUC), University of California, Berkeley (Berkeley)), as reported through each asymptomatic surveillance and diagnostic testing program [40, 41, 42]. The testing program used pooled saliva sample surveillance with follow-up diagnostic testing. The positivity rate was reported each day, but individuals must wait at least 1 week between tests. We aggregated the positivity rate by week during the Fall 2020 semester.

Confirmed case data

When calibrating our NABM, we considered the reported confirmed cases in the counties where the three campuses locate [43, 44, 45]. The 'Confirmed COVID-19 Cases' reported in this dataset are cases that have been confirmed with a positive molecular (PCR) test. We considered cases during the Fall 2020 semester to inform external transmissions in the NABM.

2.5 WIMOB SEIR NABM

In this section, we will construct an NABM that captures the spread of COVID-19 between individuals active within universities' community. The model is used to evaluate the effectiveness of different policy interventions and to validate reverse engineering algorithms for reconstructing the epidemic over time. We consider a modified version of the SEIR framework for simulating the spread of COVID-19 [46, 33] by using an underlying contact network given by WIMOB, the WiFi mobility data in GT. Figure 2.2 shows the compartments of the framework. The *susceptible* state (S) represents individuals who have not been infected and can contract the disease by having contact with an infectious individual. The *exposed* state (E) is canonically equivalent to the "incubation period" and is similar to the pre-symptomatic state found in related work [31, 32]. Individuals are considered *infectious* when they are in either the *asymptomatic* state (Asym) or *symptomatic* state (Sym). Individuals in the *asymptomatic* state are assumed to be the major "spreaders" [32] and transmit the infections to *susceptible* individuals before they are *recovered* (R) [47] — after 7 days [32]. Since *asymptomatic* is considered a state of mild severity [48], individuals in this state do not have a risk of fatality. By contrast, for individuals in the *symptomatic* state, will be eventually *isolated* (Iso) (e.g. self-quarantine, or hospitalization on campus). Once in the *isolated* state, they cannot transmit the disease to individuals in the *susceptible* state. Unlike the *asymptomatic* track, the *symptomatic* state is considered critical severity. Therefore, after moving to the *isolated* state, individuals have risk of fatality and entering the death state (D). If the *isolated* individual survives, they enter the *recovered* state. We assume immunity is preserved and therefore after recovery the individual is no longer *susceptible*.

Definitions

Let $t = \{0, 1, 2, 3, ..., T\}$ be the index of days in simulations. We denote the sequence of dynamic collocation networks indexed by day t, as $\{G_t(A_t, B_t)\}_{t=0}^{T}$. A_t is the set of vertices, i.e. individuals on campus, and B_t is the set of edges. The universe set of the population throughout the simulation time period is given by $M = \bigcup_{i=1}^{T} A_t$. For convenience, we use $a_i \in M$ to index every person in the universe population set.

The SEIR model consists of seven compartments. Each of these corresponds to a function of population subsets with respect to day t: *susceptible* S(t), *exposed* E(t), *asymptomatic Asym*(t), *symptomatic Sym*(t), *isolation* I(t), *recovered* R(t), and *dead* D(t). For example, $a_i \in I(t)$ means a_i is in the *isolation* state at day t. We use $N^t_{S\to E}$, $N^t_{E\to Asym}$, $N^t_{E\to Sym}$, $N^t_{Asym\to R}$, $N^t_{Sym\to I}$, $N^t_{I\to R}$, and $N^t_{I\to D}$ to denote the transitions between states between day t and day $t + 1$.

Model Initialization

The entire population M is fixed where $M = S(t) + E(t) + \textit{Asym}(t) + \textit{Sym}(t) + I(t) + R(t) + D(t)$ for all t. To capture the positivity out of the students coming back to campus at the start of the semester, we initialize the system by setting a subset of M into $\textit{Asym}(0)$ and the reminder into $S(0)$. The initial percentage of *asymptomatic* is described by:

$$\textit{Asym}(0) \sim \textit{Binomial}(M, I_0)$$
$$S(0) \sim M - \textit{Asym}(0)$$

where I_0 is a parameter defined as the initial percentage of *Asymptomatic* at day $t = 0$.

New exposures

We consider two ways that an individual in the NABM could be exposed: (i) exposures that occur due to contacts among individuals captured by the mobility network (*internal*

transmission) and (ii) exposures that occur due to contacts that occur outside of the mobility network (*external transmission*).

Internal transmissions happen exclusively among individuals in the model. On any given day, an edge becomes effective, when one of the *susceptible* individual comes in contact with the other which is infectious, i.e. *asymptomatic* or *symptomatic*, individual. Therefore, for every effective edge between two such people, the probability of the susceptible individual getting *exposed* is described by the transmission probability p, which is another model parameter. The probability for an susceptible individual a_i entering *exposed* at the end of day t is given by the following function:

$$f_p(a_i, t, p) = \begin{cases} 1 - (1-p)^{e(t,a_i)}, & \text{if } a_i \in V_t \\ 0, & \text{otherwise} \end{cases}$$

Here, $e(t, a_i)$ is the number of effective edges of individual a_i at time t. Since $(1-p)^{e(t,a_i)}$ is the probability that a_i does not contracted the disease at time t under $e(t, a_i)$ Bernoulli trials, $1-(1-p)^{e(t,a_i)}$ is the probability that at least one effective edge leading a_i to *exposed*.

In addition to exposure due to internal transmission, we also consider new exposure due to external transmission. We consider external transmission to be exposure resulting from the physical collocations outside the scope of mobility network. For instance, the WIMOB does not capture the connections between individuals without access to the campus WiFi or someone contacting infectious persons outside the campus. To reflect this risk in our model, for any day t, $I_{out}(t)$ describes the probability of infection on day t from a collocation that is external to the mobility network. We assume that the probability an individual is infected due to an external source is proportional to the number of cases in the broader community. Therefore, we model the probability of external infection as a function of confirmed cases in the surrounding county, e.g, Fulton county, where the Georgia Institute of Technology (GT) is located [43]. C_t represents the confirmed cases reported by Fulton County where

C_{max} is the maximum number of the cases over the whole period, $I_{out}(t)$ is given by

$$I_{out}(t) = \alpha * \frac{C_t}{C_{max}}$$

where α is a parameter scaling the normalized confirm cases in the surrounding county. The resulting number of external infections on day t is then modeled to be are Binomial with $|S(t)|$ trials with probability of success $I_{out}(t)$.

In summary, for every day $t > 0$, the overall number of individuals that become newly exposed is represented as $N^t_{S\to E}$ which is the result of both external and internal transmissions.

$$N^t_{S\to E} \sim \underbrace{\textit{Binomial}(|S(t)|, I_{out}(t))}_{\text{external infections}} + \underbrace{\sum_{a_i \in M} f_p(a_i, t, p)}_{\text{internal transmissions}}$$

Model dynamics after exposure

After exposure, individuals in the model will progress through other disease states in our model. We update the number of individuals in each state daily to reflect transitions between them. The transitions between the states on day t are summarized according to the following equations:

$$
\begin{aligned}
S(t+1) - S(t) &= -N^t_{S\to E} \\
E(t+1) - E(t) &= N^t_{S\to E} - N^t_{E\to \text{Asym}} - N^t_{E\to \text{Sym}} \\
\textit{Asym}(t+1) - \textit{Asym}(t) &= N^t_{E\to \text{Asym}} - N^t_{\text{Asym}\to R} \\
\textit{Sym}(t+1) - \textit{Sym}(t) &= N^t_{E\to \text{Sym}} - N^t_{\text{Sym}\to I} \\
I(t+1) - I(t) &= N^t_{\text{Sym}\to I} - N_{I\to D} - N_{I\to R} \\
R(t+1) - R(t) &= N_{I\to R} \\
D(t+1) - D(t) &= N_{I\to D}
\end{aligned}
$$

After an individual has been exposed, they will spend Δ_{S} days in an incubation period. At day Δ_{S} after their exposure, individuals will become a *symptomatic* infection with probability p_S. Otherwise the agent will become an *asymptomatic* infection This process is given by the following two equations:

$$
N^t_{E\to \text{Sym}} \sim \begin{cases} \textit{Binomial}(|E(t-\Delta_{\text{S}})|, p_S), & t \geq \Delta_{\text{S}} \\ 0, & \text{otherwise} \end{cases}
$$

$$
N^t_{E\to \text{Asym}} \sim \begin{cases} |E(t-\Delta_{\text{S}})| - N^t_{E\to \text{Sym}}, & t \geq \Delta_{\text{S}} \\ 0, & \text{otherwise} \end{cases}
$$

Individuals who enter the *asymptomatic* state will recover after $\Delta_{\text{Asym}\to R}$ days since they were first *exposed*. Thus, we represent the number of transitions from *asymptomatic* to *recovered* on day t as:

$$N^t_{\text{Asym}\to R} \sim \begin{cases} N^{t-\Delta_{\text{Asym}\to R}}_{E\to\text{Asym}}, & t \geq \Delta_{\text{Asym}\to R} \\ 0, & \text{otherwise} \end{cases}$$

On the other hand, individuals who enter the *symptomatic* will eventually enter the *isolation* state [32]. The time that individuals spend in the *symptomatic* state before entering the *isolated* state is normally distributed $\delta^t_I \sim \textit{Normal}(\Delta_I, \sigma^2_I)$. We simulate each individual's transition between *symptomatic* and *isolated* by using a sampling function $\Gamma(a_i, t, \Delta_t)$ and a function $\tau(a_i, t)$ that returns the days since *exposed* respectively:

$$\Gamma(a_i, t, \delta^t_I) = \begin{cases} 1, & t - \tau(a_i, t) \geq \delta^t_I \\ 0, & \text{otherwise} \end{cases}$$

$$\tau(a_i, t) = \begin{cases} \text{first day of } a_i \text{ entering } \textit{exposed}, & a_i \in \textit{Sym}(t) \\ +\infty, & \text{otherwise} \end{cases}$$

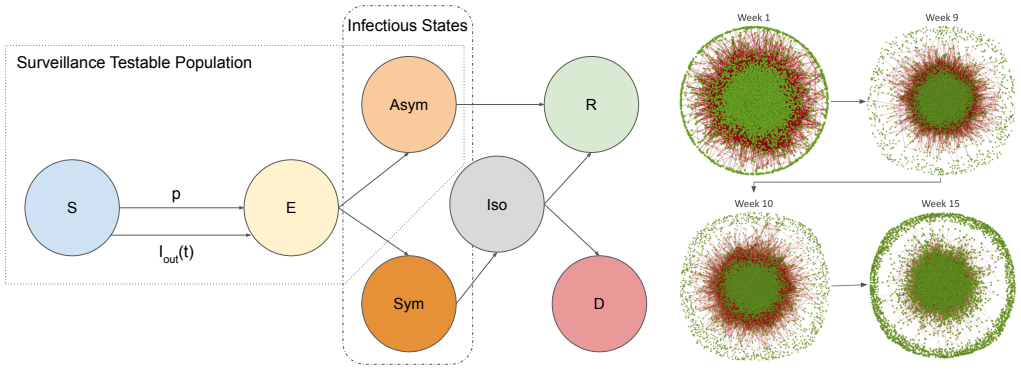
The aggregated transitions $N^t_{\text{Sym}\to I}$ between *symptomatic* and *isolated* is the sum of the distribution above on each day t.

$$N^t_{\text{Sym}\to I} \sim \sum_{a_i \in M} \Gamma(a_i, t, \delta^t_I)$$

Individuals who enter the *isolated* state may end up with one of two states: *dead* or *recovered*. We defined $N^t_{I\to D}$ as following another binomial distribution with parameter p_D:

$$N^t_{I\to D} \sim \textit{Binomial}(|I(t)|, p_D)$$

The transitions between *isolation* and *recovered* is quite similar to the transitions between



(a) (b)

Figure 2.2: (a) The schematic of the compartments in our modified SEIR model. By the design of the GT surveillance testing [40, 49], the total testable population is defined as the summation of *susceptible*, *exposed*, and *asymptomatic*. Infectious persons are in either *symptomatic* or *symptomatic*. For every effective edge in the mobility network, a susceptible individual that is exposed to an infectious person becomes infected with probability p. Individuals may also get infected due to an exposure not captured by the WIMOB network which occurs with probability $I_{out}(t)$ on day t. account for new infected cases. (b) The mobility behavior represented by WIMOB changes every day of the semester (shown weekly here). The contact network constructed from WIMOB forms the underlying contact structure of the NABM.

symptomatic and *isolation* except $\delta_R^t \sim Normal(\Delta_R, \sigma_R^2)$ where Δ_R and σ_R are the two parameters standing for the mean and standard deviation of days for an individual in the $isolation$ state entering $recovered$ since the first day of infection. This leads to:

$$N_{I \to R}^t \sim \sum_{a_i \in M} \Gamma(a_i, t, \delta_R^t).$$

2.6 Calibration

Most of our model parameters can be estimated from previous studies (see Table 2.1). However, three parameters in our study are not easily estimated from previous studies: (1) the proportion of the agents that begin the semester asymptotically infected, I_0, (2) the probability of transmission between a given infectious individual and susceptible individual given a contact in the mobility network, p, and (3) the scaling factor α used to determine

probability of transmission due to contact outside of WIMOB network on day t, $I_{out}(t)$ (see (section 2.5)). We fit these three parameters to the published weekly positivity rate (percentage of asymptomatic cases) as reported byUIUC, Berkeley, and GT's asymptomatic surveillance testing program [40, 42, 41]. To fit the parameters, we performed calibration to minimize the root mean square of error(r.m.s.e) between the simulation estimates of the weekly positivity rate and the observed weekly positivity rate on ,UIUC, Berkeley, and GT's campus of the Fall 2020 semester as reported by each surveillance testing program.

To perform the calibration, we used two sets of public data pertaining to 2020 Fall semester at each school: (i) the confirmed cases in the school's counties [43, 45, 44], and (ii) the aggregated surveillance test positivity rate for each week [40, 41, 42]. The former helps estimate the daily external infection percentage. The latter is the ground truth trajectory we fit our model on. We consider the data aggregated by week because each individual on campus can only get tested once per week. The positivity rate provided by the surveillance testing data can be interpreted as the estimated percentage of new *asymptomatic* cases out of the total testable population which includes *susceptible*, *exposed*, and *asymptomatic* — with an assumption that every testable population get tested at the same rate.

To formalize the calibration problem, let R_w be the surveillance-testing aggregated result at week w. Let $S(I_0, \alpha, p, w)$ be the function of the simulation model which returns the percentage of new *asymptomatic* in week w out of the total testable population. For every combination of parameters, the predicted result for each week w is estimated by taking the average of N simulation outputs. The objective function is:

$$f(I_0, \alpha, p) = \sqrt{\frac{1}{W}\sum_{w=1}^{W}\left(\frac{\sum_{i=1}^{N} S(I_0\alpha, p, w)}{N} - R_w\right)^2}$$

The optimization problem is:

$$\min_{I_0,\alpha,p} f(I_0, \alpha, p)$$

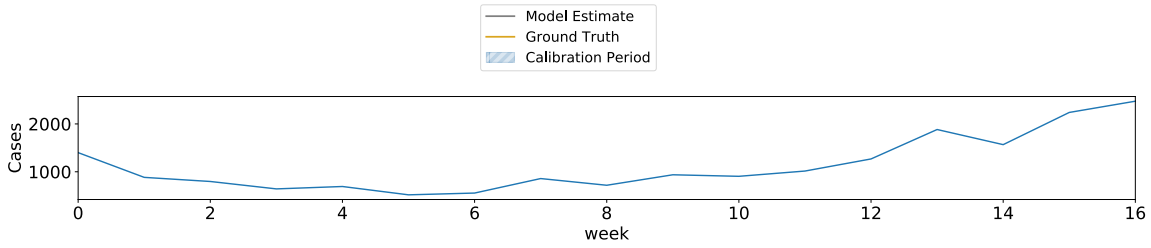
Table 2.1: Model Parameters of the NABM on GT

Parameter	Definition	Value	Std	Source
p	Transmission probability: For any edge between a *susceptible* and *infectious* individual in the contact network, p is the probability that the *susceptible* person will enter into the *exposed* state. This only dictates internal transmission	0.034	0.007	Calibration
α	Scaling factor of the normalized confirmed cases in the surrounding county((section 2.5)). This is the parameter for us to generate $I_{out}(t)$	0.032	0.0032	Calibration
I_0	Proportion of population that is *asymptomatic* at day 0	0.012	0.0009	Calibration
p_S	Probability of *exposed* persons becoming symptomatic	0.66	-	[32]
Δ_S	Incubation period (days) since the first day of exposure	5	-	[32]
$\Delta_{\text{Asym}\to R}$	Asymptomatic duration (days); it is the time taken for an *asymptomatic* person to recover since the first day of exposure	7	-	[32]
Δ_I, σ_I	Time of an *symptomatic* entering *isolated* since the first day of exposure of a *symptomatic* person	8	2	[51]
Δ_R, σ_R	Time for recovery for a *symptomatic*, since the first day of exposure	12	2	[10]
p_D	Death rate under *isolated*	0.0006	-	[10]

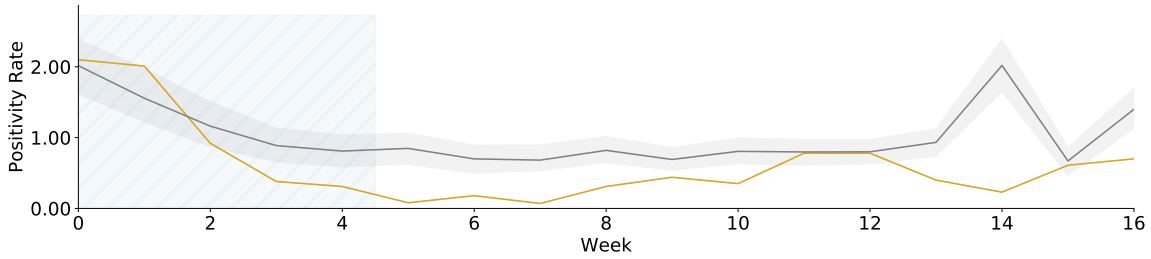
We fit our model to the first 5 weeks of Fall 2020 and validate the results on the remaining weeks. After obtaining the optimal set of parameters, for robust comparison of policies with different viral variants, we generate a range of parameters by compromising the r.m.s.e within 40% of the minima [33]. First, we implement the Nelder Mead method [50] to discover the optimal set of parameters that minimizes the r.m.s.e. Next, we sample 40 different combinations of parameters within 40% of the minimum r.m.s.e to estimate the means and standard deviations of these parameters (Table 2.1). Throughout this paper, we pool together all simulation results across those parameters over multiple runs ($N = 15$) and report the 2.5th and 97.5th percentiles of the simulation outputs for every policy experiment. For parameters other than GT, please check in Table A.1 in the appendix chapter.

2.7 Validation

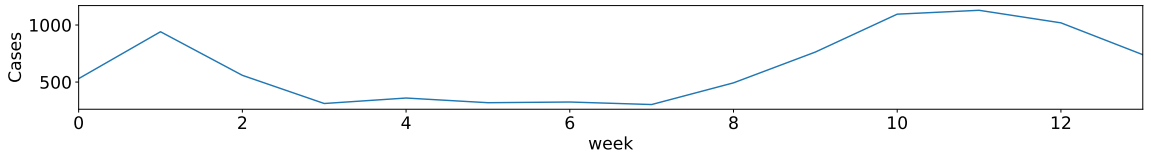
We split the available dataset in the Fall semester of 2020 into a training set described in the previous section (the first five weeks since 08/17/2020) and a test set (the remaining 12 weeks). We then validate the model by comparing simulation results with the surveillance data reported not used for calibration in different campuses in the Fall semester of 2020, the asymptomatic percentages of the total testable population. Note that the evaluated NABMs use only the optimal parameter set (not the 40% approximation set for quantifying uncertainty). We keep running the models after the fifth week of the semester. As seen in Figure 2.3, for the NABM in every campus, the simulation results after the first five weeks follow patterns fairly close to the ground truth with a 2-3 weeks delay and overestimation in general.



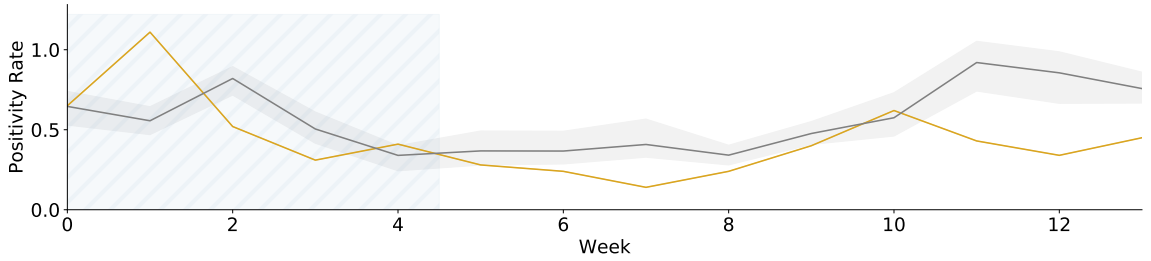
(a) External cases: GT



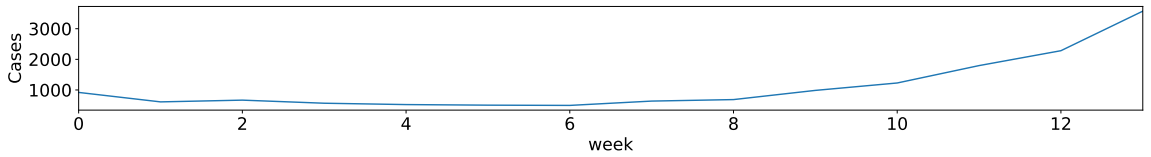
(b) Calibrating on the weeks 0-4: GT



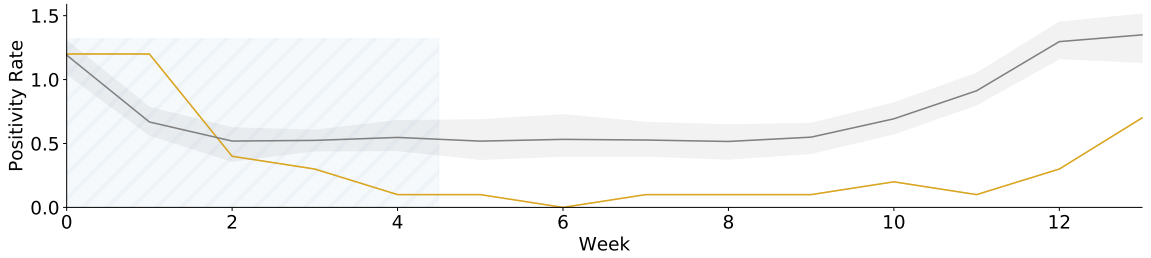
(c) External cases: UIUC



(d) Calibrating on the weeks 0-4: UIUC



(e) External cases: Berkeley



(f) Calibrating on the weeks 0-4: Berkeley

Figure 2.3: Calibration and Validation Results

CHAPTER 3
POLICY DESIGN AND EVALUATION

3.1 Overview

Infectious diseases can cause devastating impacts on university campuses. The respiratory infectious disease that appeared in 2020, Covid-19, is indeed challenging universities in the world. In the U.S., more than half a million cases have been identified at universities [3]. As the global pandemic continuously spreading throughout the country, it is common for school officials to adapt non-pharmaceutical interventions (NPI) such as the closure of locations to control on-campus infections [52]. These operation strategies typically rely on remote instruction (RI) to maintain educational activities without using campus-spaces [53, 54, 55]. Recent studies tell us that policy-makers can leverage course-enrollment (Course enrollment data (EN)) data to design RI policies: move classes with many attendants to online mode [19, 17]. Since the national lockdown, 44% of the colleges and universities in the U.S. have implemented RI [56]. Nevertheless, these policies have side effects by forcing too many students into a completely remote schedule [57]. For example, students' learning outcomes might be worse due to losing months of education, which might affect their future careers [58]. Besides, the economy of universities and their surrounding neighborhood might go down if the campus population is small [6, 7].Furthermore, with socioeconomic disparities and heterogeneous household contexts, the demands of remote instruction can lead to added anxiety and stress among students [59, 60]. Therefore, relying on RI is an inappropriate approach for the universities to balance community health with the demand for education, social economy, and wellbeing. A more versatile and robust approach is in need such that policy-makers have a good assessment of closure's side-effect while conceiving policies for minimizing the risk of future outbreaks.

3.2 Contribution

We propose a new method to design and experiment with novel closure policies in a university campus setting by utilizing WiMob from the campus's WiFi infrastructure. First, WiMob involves constructing anonymized mobility networks that inform an extended period of collocation or "proximate contact" between individuals to depict the real contact networks. One predominant nature of WiMob is that it enables closure policies at the granularity level of rooms, halls, or suites. We define our closure policies equipped with the WiFi-based mobility data as localized closure LC. Second, WiMob utilizes the data collected over 2019 and 2020 of approximately 40,000 anonymous occupants and visitors of the Georgia Institute of Technology, a large urban campus in the U.S. - including about 16,000 undergraduate students, 9,000 graduate students, and 7,600 staff members. Third, WiMob is further used to design and analysis of LC. We evaluate the performance of LC in comparison to RI based on an NABM model with WiMob as the contact-networks among agents. The model is calibrated with GT on-campus Covid-19 cases from the fall semester of 2020 and Fulton County's infection rates. To compare the effectiveness of interventions, we describe a counterfactual semester unaltered by other policy-induced behaviors of 2020 by leveraging WiMob from fall 2019 to determine the simulation's contact structure. This study's designed experiments aim to assess RI and LC's performance metrics related to disease size and burdens under different behavioral scenarios.

As a result, we find that LC is comparable to RI in controlling total infections but more effective at reducing peak infections and internal transmission. Concerning the burdens to campus and individuals, LC targets fewer locations and forces fewer students into fully online mode, and does not isolate any more people than IC. In general, WIMOB can assist universities in devising highly-specific closure policies, like LC, which can simultaneously contain the spread of disease and mitigate campus disruption compared to RI.

3.3 Policy Design

RI: Offering Large Classes Online

As a response to Covid-19, prior work has recommended using EN to enforce a form of RI— moving classes large to an online remote instruction setup while other classes are offered in–person [18, 19, 17]. While we have access to aggregate insights on EN contact networks, our study protocol prohibits us from accessing course-specific information at an individual level. Therefore to infer individual enrollment, we analyze the edges of the bipartite graph K. For this, we first scrape the GT's course roster for Fall 2019 (filtered to only represent the Atlanta campus). This process provides us with a location and weekly schedule for every lecture conducted on campus, including its various sections. With this information, we are able to identify which edges represent visits to lectures, and subsequently, we can account for unique visitors to a lecture. Thus, we can first identify the number of unique individuals on campus who are enrolled in classes. The aggregate data from course enrollment reports that $21,299$ students were enrolled in Fall 2019. In comparison, our inference identifies $22,248$ students. The excess number can be explained by the fact that our method does not distinguish between instructors, TAs, and students. Next, we study the unique visitors to every lecture in the scraped course schedule which gives us an estimate for the size of every class. Given the limitations of our data processing, actual enrollment sizes could be larger, but our process is less likely to count false positives [38]. Finally, to model RI, for the contact network G_t, we create a counterfactual network G'_t for each day t. These exclude collocations that took place at lecture locations during lecture times. If two people were connected solely by proximity during lectures — in a class with large enrollment — they will appear disconnected in the counterfactual network.

LC: Closing Important Locations

This article demonstrates the effectiveness of localized closures,LC, which are targeted interventions to seize mobility at different spaces on campus. For this, we identify important locations on campus by analyzing H. In the main paper, LC uses *PageRank* [61] as an illustrative algorithm to identify important location nodes. For robustness, we apply various additional algorithms to identify highly authoritative nodes in H — betweenness centrality [62], eigenvector centrality [63], and load centrality [64]. In the SI Appendix, we distinguish these different policies as Localized closure policies with Pagerank ($\mathsf{LC}_{\text{PRank}}$), Localized closure policies with Betweenness Centrality ($\mathsf{LC}_{\text{BCen}}$), Localized closure policies with Eigenvector Centrality ($\mathsf{LC}_{\text{ECen}}$), Localized closure policies with Load Centrality ($\mathsf{LC}_{\text{LCen}}$). Since RI captures a weekly schedule to determine enrollment, LC is implemented to find locations based on behavior from the past 7 days of mobility. We apply the weighted version of the algorithms mentioned earlier on the directed graph representing movement, H. The edge weight is based on the number of instances of movement between any L_i and L_j. After sorting the locations by importance, we determine the number of locations to shut down based on different budgets induced by RI— mobility and risk of exposure. For this purpose, we take the approach of a greedy algorithm which successively removes highly-ranked locations till the constraint is met (within 1% margin of error). Similar to RI, LC also render counterfactual collocation networks, $G"_t$ for each day t. In these networks, we remove instances of collocations that occurred at the shutdown locations.

3.4 Inducing Budgets and Characterizing Behavioral Scenarios

We now describe how we compare the RI and LC policies. First, we consider the effects of these policies under three behavioral scenarios. These scenarios express the spillover effects of closure that lead to students avoiding campus entirely because their entire schedule is forced online. This analysis assumes that the motivation to be present on campus

is determined primarily by enrollment. We consider that, if a student has a full course load (enrolled in a minimum of 3 classes) and all their classes are offered online, that student might have less incentive to visit campus at all (for any engagement) and thus practice *Avoidance*. Since LC could end up closing classrooms, it can also lead to academic schedules being affected and elicit *Avoidance* behavior. As a result, we describe three scenarios. Persistent Scenario (*S1*), is the preliminary, or null scenario, which represents no *Avoidance*. This counterfactual collocation graph only removes edges directly affected by RI or LC. The second scenario we model is Non-residential Avoidance (*S2*) where only non-residential students with full online schedules stop visiting campus entirely. Here the counterfactual graph will remove all edges of non-residential students with fully online schedules. Lastly, the third scenario we model is Complete Avoidance (*S3*) where any student with fully online schedules stops activity on campus entirely (including residential students). Here the counterfactual graph will remove all edges from any student with fully online schedules. Since our study protocol prohibits us from mapping our data to other sources, we heuristically infer which individuals are likely to be residential and which are not. We label individuals as residential when they dwell an average of at least 15 minutes at residential locations between 6pm and 10am, on workdays (Monday–Thursday).

Under each scenario, we limit the number of locations that can be closed under the LC policy to ensure the level of restriction is constrained to be similar to the RI policy. We limit the number of locations under two types of restrictive budgets. The first budget is based on *mobility*, which is the percentage of edges remaining in the bipartite graph if a policy were to be implemented. The second budget is based on *exposure risk*, which is the number of unique individuals who would be in the 1-hop collocation neighborhood of positive individuals. We compute this budget by randomly sampling 2.5% of the population as positive, based on the highest 7-day average positivity rate reported by GT [49] in Fall 2019. Note, however, the effect of RI on campus can vary in different behavioral scenarios, thereby changing the budget available to design a comparable LC policy. For instance, the number

of people at exposure risk is much lower in *S3*. As a result, we build multiple alternate networks representing the effect of policies under counterfactual behavioral scenarios.

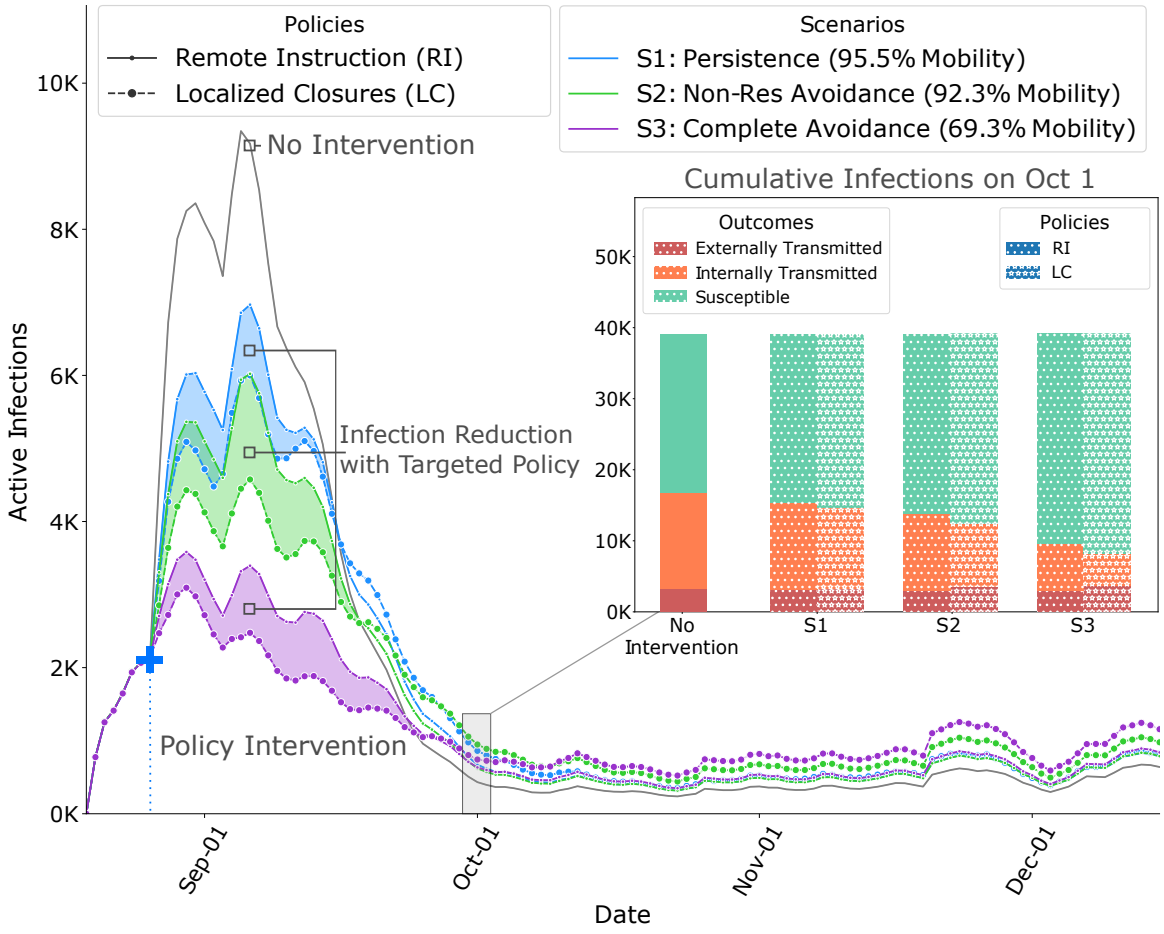
3.5 Outcome Metrics

We present differences between LC and RI based on three infection reduction outcomes; peak infections (maximum active cases on a given day), internal transmission (exposure from infected individuals on campus), and total infections (cumulative cases at the end of the semester). Additionally, we measure the burden of policy interventions with the number of locations closed — requires resources to monitor and maintain super-spreader locations, the percentage of students that avoid campus — disruption to learning outcomes [58, 59], and the percentage of individuals completely isolated — worsens mental wellbeing [65].

3.6 Results

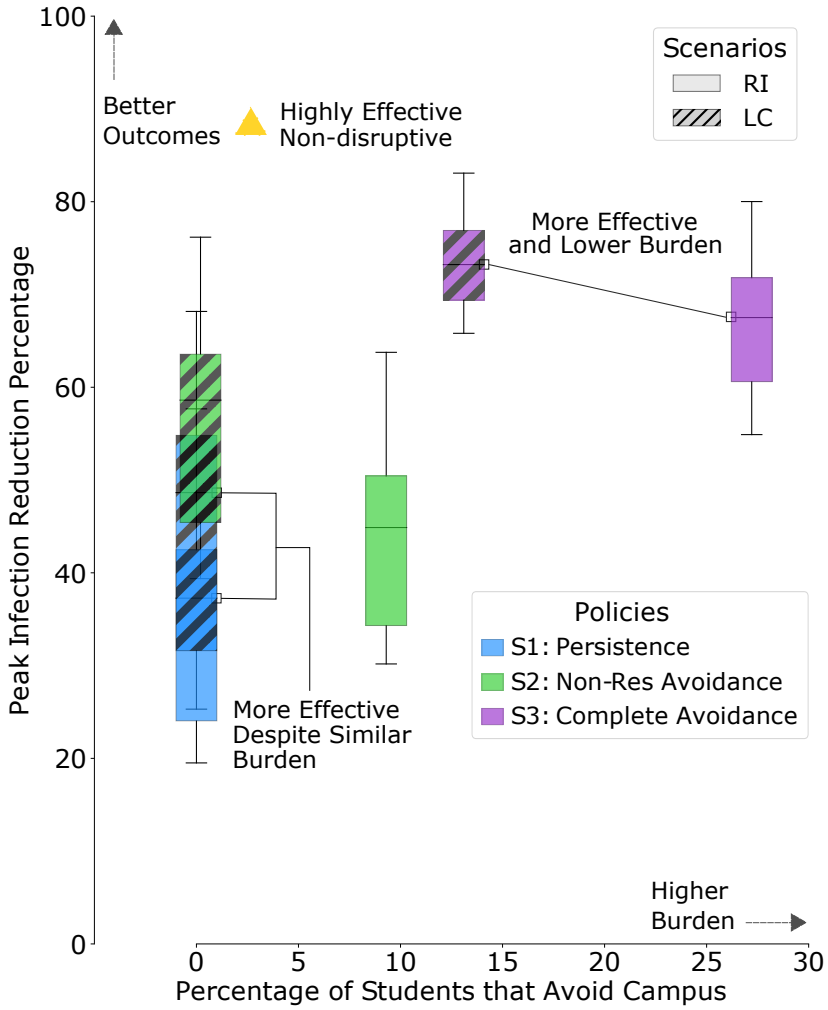
LC cause greater reduction in peak infections, while affecting fewer locations

Controlling peak infections relaxes the burden on a university to support positive cases for any given day, and allows resources to be distributed over time. In all scenarios, of our simulation of Fall 2019, we observe that the peak reduction is significantly better in LC (Table 3.1) than RI. While RI impacts 58 different locations (classrooms and lecture halls), in *S1* and *S2*, LC achieves better outcomes by closing fewer locations. For example, in *S2*, RI achieves a 28.9% peak reduction, but LC shows a reductions of 49.3% (mobility budget) and 48.1% (exposure risk budget). This is attained by closing 38 or 50 locations respectively. Therefore, with such policies, policymakers need to restrict fewer locations to remarkably minimize the pressure of active infections on campus (e.g., diagnoses, treatment, quarantining).



(a)

Figure 3.1: Results of policy interventions with our calibrated NABM on contact networks from Fall 2019, derived from WIMOB. This graph compares the mean active infections between LC and RI. LC show improved outcomes (shaded regions) even when constrained to the same restrictions of RI policies. (a)–*inset:* After the first wave, even though LC shows slightly higher active infections, the cumulative infections are still lower, especially those that are a result of internal transmission on campus.



(a)

Figure 3.2: Outcomes of policies within the same scenario are shown with boxes of the same color (RI policies are solid, LC policies are hatched) and box heights represent the 2.5^{th} and 97.5^{th} percentile. In *S1*, even though LC and RI are equally burdensome in terms of students avoiding campus, LC shows improved outcome on peak reductions. In fact, for the other scenarios, LC shows better outcomes than RI, without forcing as many students into online schedules, and, therefore, being even less burdensome with greater impact.

Table 3.1: Comparison of different policies in terms of controlling the disease and impacts on campus in Fall 2019. Within each scenario, we perform the Kruskal-Wallis H-Test [66] to compare outcomes of LC with RI. We find that LC leads to significantly improved peak infection reduction and internal transmission. In terms of reduction in total infections, the outcomes are comparable in general but can vary by specific scenarios. In addition, every policy also exerts some burden on campus, either in terms of locations affected, students avoiding campus or isolation. We observe that LC policies focus on fewer locations (except in (except in *S3*)). Moreover, these policies affect fewer student's schedules and therefore fewer people avoid campus due to completely remote schedules. Finally, LC does not increase the percentage of people completely isolated on campus (p-value: < 0.01:*, < 0.001:**).

Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	RI	LC		RI	LC		RI	LC	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	$25.34(\pm12)$	$36.92(\pm14)^{**}$	$34.30(\pm13)^{**}$	$35.44(\pm10)$	$49.33(\pm11)^{**}$	$52.19(\pm10)^{**}$	$61.62(\pm7)$	$69.34(\pm5)^{**}$	$64.44(\pm6)^{**}$
Total Infections (%)	$6.99(\pm5)$	$10.63(\pm6)^{**}$	$8.19(\pm5)^{**}$	$14.88(\pm4)$	$13.96(\pm6)^{*}$	$15.67(\pm6)$	$33.00(\pm5)$	$33.4(\pm5)$	$26.94(\pm5)^{**}$
Internal Transmissions (%)	$17.13(\pm9)$	$22.62(\pm11)^{**}$	$21.01(\pm11)^{**}$	$27.58(\pm8)$	$35.35(\pm12)^{**}$	$39.20(\pm11)^{**}$	$54.00(\pm8)$	$70.89(\pm7)^{**}$	$60.90(\pm9)^{**}$
Burdens on Campus									
Locations Affected	58	18	19	58	38	50	58	192	124
Students Avoiding (%)	0	0	0	9.30	0.20	0.45	27.21	12.45	6.57
Completely Isolated on Campus (%)	5.42	8.40	8.40	5.95	5.72	5.71	7.09	5.18	5.23

Table 3.2: Comparison of Contact Network Structure (Fall 2019). We create a contact network of only students with WiMob and compare it with insights from contact networks created with En. On average, we find the contact network constructed with WiMob shows fewer average contacts, lower density and higher average shortest path (between reachable paths). Moreover, within WiMob itself, characterizing all spaces reveals more contacts and shorter paths than only focusing on contacts in lectures. While the proportion of the largest component appears similar, note that with WiMob, on average about only 70% of the students visit campus on a given week. We further inspect the disease–mitigating structural changes of the RI policy on the network. We observe that the changes across all metrics with En appear to be more drastic than compared to WiMob.

	Cornell		Georgia Tech				
Contact Network	En		En		WiMob		
Contact Situations	Course Lectures	RI	Course Lectures	RI	All Spaces	Course Lectures	RI
Number of Active Nodes	22051		21299		15379($\pm$3353)	15379($\pm$3353)	15380($\pm$3353)
Average Contacts	529	22 – 41	341	30	152($\pm$63)	86($\pm$35)	86($\pm$34)
Density	0.024	0.001	0.016	0.001	0.009($\pm$0.002)	0.005($\pm$0.001)	0.0053($\pm$0.0014)
Largest Connected Component(%)	0.991	0.763	0.994	0.627	0.999($\pm$0.001)	0.999($\pm$0.02)	0.978($\pm$0.025)
Average Shortest Path	2.47	3.75	2.54	3.54	2.67($\pm$0.28)	3.26($\pm$0.5)	2.953($\pm$0.35)

LC lead to comparable reduction in total infections, while keeping more students on campus

Universities want to minimize the number of infected cases while ensuring majority of the population remains active on campus to continue successful operation. The total number of infections reduced by both LC and RI are similar. While the differences between policies are statistically significant (Table 3.2) in some scenarios, the magnitude of these differences might not be practically as important. In contrast, the impact the policies have on the student schedules is remarkably different. RI forces multiple students to adapt to fully online schedules. In the Scenario *S2*, 9% of students do not visit campus and in *S3*, 27% of students do not visit campus. On the other hand in LC the number of students expected to avoid campus can be as low as 0 and never exceeds more than 12%. Besides sustaining economic loss to the campus, remote instruction can increase anxiety among students and hinder learning outcomes [59, 60]. Compared to RI, LC offers policymakers a way to defend against turnover in the student population, without compromising overall control of disease spread (Table 3.1). Limiting the number of students that avoid campus helps preserve on-campus businesses [6, 7] and minimally disrupts the student wellbeing.

LC cause greater reduction in internal transmission without causing further isolation on campus

Universities are responsible for limiting spread on campus, but they must also ensure that aggressive policies do not worsen mental wellbeing of the community. In terms of internal transmission the reduction is significantly larger with LC (Table 3.1). However, when LC restricts the infections early in Fall 2019, it leaves more individuals susceptible to external transmission. College student behavior outside campus on weekends and breaks is known to impact local transmission [67]. When policymakers consider LC they should also consider policies on re-entry or required testing based on off-campus activities. In terms of isolating individuals on campus, it's notable that LC and RI are similar in *S2*. Interestingly,

in *S3*, where LC closes more than 100 locations, the percentage of isolated individuals per week is less than that of RI. This finding implies that LC can keep individuals on campus without forcing them into complete isolation. Here "isolation" refers to no form of proximate contact with any individual on campus — extreme social distancing where individuals are not even collocated in the same suite or hall. While social distancing is a recommended countermeasure for COVID-19 [54], complete isolation can have adverse effects on psychological wellbeing [65, 68, 69]. Staying completely isolated on campus can increase loneliness and limit social connectedness [68], which are both related to depression [65]. Although the proportions are similar (Table 3.1), LC does not necessarily isolate the same sets of individuals. This qualitative difference could also explain the difference in internal transmissions — LC could be isolating individuals who are less likely to spread the virus.

LC identifies a wider variety of auxiliary spaces.

By using WIMOB to design LC we are able to identify locations for closure at the granularity level of rooms, including unbound spaces such as lobbies and work areas. First, in *S1*, we find that most locations that LC targets are a subset of the auditoriums–like rooms where large classes would take place in Fall 2019. Note, LC needs to restrict only a few such spaces to be under the same budget as RI. This is because, under *S1*, RI policies only alter visits to lectures, while these spaces are used for other purposes during other times (e.g., club activities and seminars). We also note that LC targets 'high traffic' locations like conference center lobbies which are typically used as waiting areas or for networking events. Next, in Scenario *S2*, we see that in addition to spaces mentioned earlier, interestingly LC further restricts the use of smaller rooms (occupancy $13 - 35$) which would not be affected by RI (as only classes of size ≥ 30 are offered online). LC also targets areas in the recreation center (which includes locker rooms and indoor courts for $4 - 20$ people). This insight indicates that our methodology WIMOB is sensitive to other student activities. Moreover, we also find a selection of spaces that would not be frequented by the undergrad-

uate population, such as lab areas and facility buildings like the police station. Lastly, in Scenario *S3*, LC targets closure of activity in far more spaces than RI. However, the better outcomes can be attributed to the fact that LC diversifies the potential restriction areas. LC now restricts heavily used small study rooms or breakout rooms (for $1-6$ people). Furthermore, it restricts use of spaces where multiple small groups of people can organically assemble, such as cafes, dining halls, and reading areas. We also observe that LC restricts activity in about 10 Greek Houses but does not target other housing areas — demonstrating its ability to restrict social behavior that could amplify disease spread. Figure A.17 shows the diversity in locations for various LC policies.

3.7 Sensitivity Analysis

In this section, we design complementary experiments to inspect the robustness LC policies under different setups and calibration approaches. These variations are defined as follows:

- Calibration periods (V1): For the results in the main paper, we discuss results with our NABM calibrated on the first 5 weeks of surveillance testing data. For additional analyses, the model parameters are re-estimated based on the surveillance data from week $5-9$ and $10-14$ in Fall 2020 at GT. The calibration is validated on the remaining weeks in the semester. Figure A.1 shows the calibration and validation. The results of policy comparison with these variations can be found in Table A.6 and Table A.7, for weeks $5-9$ and $10-14$ respectively. Additionally, Figure A.7 shows boxplots to compare the distributions of different policies, while Figure A.13 and Figure A.14 show cumulative plots of the disease control outcomes, for weeks $5-9$ and $10-14$ respectively.

- Campuses and counties (V2): For the results in the main paper, the calibration of our NABM reflects certain latent factors inherent to GT that could affect both mobility

behavior as well as testing results. To complement this we consider calibrating our data under different settings informed by surveillance testing from other similar large universities. This analysis is intended to represent the GT community in a different geographic setting, which is influenced by a different surrounding community, policies and resources. The new parameters are estimated based on the first 5 weeks of surveillance testing from the University of Illinois at Urbana-Champaign (UIUC) and the University of California, Berkeley (Berkeley) [41, 42], and the corresponding county data uidphcv19, acdphcv19 The calibration is validated on the remaining weeks in the semester. Figure A.2 and Figure A.3 show the calibration and validation for UIUC and Berkeley respectively. The results of policy comparison with these variations can be found in Table A.8 and Table A.9. Additionally, Figure A.8 shows boxplots to compare the distributions of different policies, while Figure A.15 and Figure A.16 show cumulative plots of the disease control outcomes.

The estimated parameters with these calibration variations are described in Table A.1. Both RI and LC are evaluated in the same infection reduction metrics and burden metrics again under scenarios S1, S2, and S3. Since the budgets are structural (mobility, and exposure risk) the LC policies are unchanged among the variants. Moreover, since the burden metrics are structural, those results are invariant.

CHAPTER 4

RECONSTRUCT AN EPIDEMIC OVER TIME

4.1 Overview

Universities might want to know who the outbreak sources are and how the disease transmits when the disease spread out on campus for some time. In particular, they are interested in reconstructing the epidemic by discovering the sources and flows causing the outbreak among individuals on university campuses. By doing so, universities can identify groups of individuals and locations at risk, they can prioritize testing for people, or they can discover people who have been likely infected but not captured by testing.

Surprisingly, we found that no current studies have targeted this question on university campuses specifically. Methodologically, identifying the sources and chains of transmissions from an epidemiological network falls into reverse engineering problems with existing efficient algorithms for solving them [70, 71, 72, 21]. However, the problem formulations in these studies typically have rigorous requirements on the input data and infection model. For example, they might expect perfect testing on individuals without false-positive cases and missing infections [70, 71]. Some of them assume the infection models are a susceptible-infections (SI) model or an independent cascade (IC) model [72, 71]. Moreover, an epidemiological network is already given to us as well.

Our main challenges to address are two: (1) We need to construct a network that can approximate the real-world contact network among the population on campus. (2) We should start with a formulation that works for the data available from real-world campuses. First, related studies highlight the use of course enrollment data to proximate network relations among individuals [19, 17]. Still, it inherently has limitations describing the population dynamics because it excludes the outside-class room's activities. Instead, we propose us-

ing the collocations network from WIMOB as demonstrated in Figure 2.1. Second, we are prohibited from using data reporting which specific individual is positive due to privacy concerns. The lowest granularity level data available to us is the number of positivities within a group of individuals which can be noisy in general [49]. Therefore, the expected problem formulation will need to reflect this uncertainty over the group-level data.

4.2 Contributions

We formulate the problem of reconstructing the epidemic given WIMOB and the report of positivity among groups of individuals to discover the sources and flows causing the outbreak on university campuses. The formulation leverages an existing work, and the inputs are similar to the real-world data available to us: (1) a time-varying collocation network and (2) a noised surveillance data reporting the number of positive cases for groups. In particular, we introduce a sampling procedure to convert the group's data into individual-based reports so that our problem can solved by calling an existing algorithm multiple times. Then, we validate the entire framework by instances from the calibrated NABM we have designed in previous chapters. Last, we present the result of a case study using the real-world data at GT in Fall 2020 semester under the Covid-19 pandemic.

4.3 Preliminaries

Let F be the set of global nodes. Let $G = (V, E)$ to be a temporal network [73]. A temporal network consists of a set of nodes $V = \{(u, t)|u \in F, t \in N\}$ where each (u, t) is a time-stamped copy of $u \in F$. A temporal network also consists of a set of edges $E = \{(u, v, t)|u, v \in F, t \in N\}$ where each (u, v, t) means there is a edge pointing from u to v at day t. Let $w : E \to R_+$ be the weight of edges E. A temporal path **P** between two time-stamps nodes (u, t_s) and (v, t_e) is a sequence of node-disjoint edges (u, w_1, t_1), (w_1, w_2, t_2), (w_2, w_3, t_3) ... (w_{j-1}, v, t_j) such that $t_s \leq t_1$, $t_j \leq t_e$, and $t_i \leq t_{i+1}$ for all $1 \leq i \leq j - 1$.

We use the term R to denote a individual-based report data in a schema of $\{u, t\}$, which is a subset of V. Also let J_1, J_1, ...J_k denote groups (subsets) of F. Let $D = \{J_i, c_i, t\}$ for all $1 \leq i \leq k$, and $t \in N$ be the group-based report data, where $c_{i,t}$ is the number of positivity found in group J_i at time t. The latest time-stamp is denoted by T.

Let $P = \{\mathbf{P}_1, \mathbf{P}_2, ...\mathbf{P}_l\}$ be a set of temporal paths. We say that P spans a subset of V, denoted by X, if and only if every $(u, t^{'}) \in X$ comes after corresponding (u, t) covered in P, i.e. $t^{'} \geq t$.

Given a set of temporal paths P, a node $u \in F$ is called a seed if there exists a (u, t) covered by P, and there is no edges $(v, u, t) \in \mathbf{P}_i$ for any $\mathbf{P}_i \in P$. The set of seeds of a set of temporal path P is denoted by $S(P)$.

4.4 Problem Formulation

4.4.1 Literature Review

The collocation network collected from the WiFi infrastructure is a temporal network. We first give a review of prior works for solving problems similar to our purpose. Shah and Zaman formalize the notion of rumor centrality for discovering a single source causing an epidemic under a SI model with a practical algorithm for attaining optimal solutions for d-level trees [70]. Lappas et al. formulate the problem of identifying k seed nodes, or effectors of a partially activated network, in steady-state under the IC model [71]. Both of these two formulations have strict assumptions over the number of seeds to be identified. Aditya and Sundareisan present another two formulations in terms of Minimum Description Length under the SI model, which can automatically determine the number of sources in the solution and allows false-positive reports within the current infection snapshot [72, 74]. However, these problem formulations do not fit our expectations because the given networks are assumed to be constant, and the hypothetical model is different than our NABM.

4.4.2 Cult Formulation

There is one formulation for the problem with assuming a temporal network and an arbitrary infection model [75]. We present the preliminary formulation of the problem as follows:

Problem TempSteinerTree: Given a temporal network $G = (V, E)$, an individual-based report R, a candidate set C, which is a subset of F, and a source size k, find a set of temporal paths P in order to minimize the sum of distance $\sum_{e \in P} w(e)$. subject to:
(i) P spans R
(ii) $S(P) \subset C$
(iii) $|S(P)| \leq k$

The formulation adapts the idea of "shorter is more likely" so that it does not requires us to model the likelihood of the temporal paths based on the infection models. Besides, R can include missing infections. The output forest will automatically discover likely missing infections within the infection cascade from the input data. Polina et al. show that an instance of TempSteinerTree can be transformed into a instance of Minimum Directed Steiner Tree problem [76], a well-known NP-hard graph theory problem in history [75].

4.4.3 Sampling Procedure

However, as said we are prohibited from using data reporting which specific individual is positive due to privacy concerns from IRB. The data available to us is D, the number of positive cases in each group at each time-stamp. We consider using a sampling procedure $f : D \rightarrow R$ by converting D into an instance of R. The sampling produce f for single group i's report D_i at any time t is defined in detail as follows:

$$f(D_i, t) = \textit{sample}(J_{i,t}, c_{i,t})$$

$$J_{i,t+1} = J_{i,t} \setminus f(D_i, t)$$

We set initialization $J_{i,0} = J_i$. Note that *sample*(X, n) is a function that randomly selects an element by a uniform distribution in each step and eventually returns a subset $H \subset X$, such that $|H| = n$. The recurrence within f indicates a sampling procedure without replacement for each group J_i.

4.4.4 Group-level-Influence

By introducing the sample procedure, we are able to formulate our problem and name it as Group-level-influence.

Problem: Group-level-influence Given a temporal network $G = (V, E)$, an group-based report $R = f(D)$, a candidate set C, which is a subset of F, and a source size k, find a set of temporal paths P in order to minimize the expected sum of distance $\mathbf{E}(\sum_{e \in P} w(e)|R)$. subject to:

(i) P spans R

(ii) $S(P) \subset C$

(iii) $|S(P)| \leq k$

4.5 Solving Group-level-influence

Solving Group-level-influence involves solving the TempSteinerTree problem. The key idea for solving TempSteinerTree is to transform the temporal network into a static network and apply the best approximation algorithm for Minimum Directed Steiner Tree [75]. Charikar et al. designed an $p(p-1)s^{\frac{1}{p}}$ approximation algorithm which recursively constructs a p-level trees by merging sub-trees with the lowest marginal normalized distance [76]. In particular for an $p = 2$ case, the transformation for the temporal network

$G = (V, E)$ allows us to accelerate the algorithm by reducing the solution space from the shortest paths between all pairs of nodes to global shortest paths between C and a single time stamped copy (u, t) for each $u \in F$. Here we present the accelerated version of the algorithm, named Cult, in Algorithm 1

Algorithm 1 Cult

procedure CULT($G(V, E), F, C, R, \alpha$)
  for $z(u) \in C$ and $v \in F$ **do**
    Calculate global shortest path between the dummy node $z(u)$ and v.
    Distance: $L(z(u), v)$, Paths: $q(z(u), v)$
  $T = \phi$
  $X \leftarrow$ the nodes in F covered by R
  while $T \neq \phi$ **do**
    $T_1^{'} = \phi$
    for $z(u) \in C$ **do**
      $T_1 = \phi$, $cost = \alpha$, $profit = 0$
      for $v \in X$ in an ascending order of $L(z(u), v)$ **do**
        $T_1 = T_1 \cup q(z(u), v)$
        $cost+ = L(z(u), v)$
        $profit+ = 1$
        if $len(T_1^{'}) > \frac{cost}{profit}$ **then**
          $T_1^{'} = T_1$
          $len(T_1^{'}) = \frac{cost}{profit}$
    $T = T \cup T_1^{'}$
    $X \setminus$ nodes covered by $T_1^{'}$
  return T

α is a parameter for controlling the size of the source $S(T)$. Note that we omit the calculation for the global shortest paths and the construction of dummy nodes for each candidate $u \in C$. Please refer to the original paper for the pseudo code in detail [75]. Algorithm 1's time complexity is $O(|C||E|)$.

Proposition 1: We can attain a solution for Group-level-influence by calling Cult W times and each time there is a new sample from $R_i = f(D)$. We then merge the results $T_1, T_2..., T_W$ to have a new forest T^*.

We propose the algorithm, Cult-sampling, based on **Proposition 1**. The pseudo code is

presented in Algorithm 2.

Algorithm 2 Cult-sampling

procedure CULT($G(V, E), F, C, D, \alpha, W$)
    $T^* = \phi$
    for $i = 1, 2, 3...W$ **do**
        $R_i = f(D)$
        $T_i = \text{Cult}(G(V, E), F, F, R_i, \alpha)$
        $T^* = \text{Merge}(T^*, T_i)$
    return T^*

In particular, we set the input candidate set C to be the global nodes F (i.e. everyone is a source candidate). Currently, we have not finalized the Merge() function yet for aggregating the forests from each sampling to form a new forest. We propose using dominator-tree [77, 78], which will be presented in our future publications.

4.6 Validation

In this section, we first validate Cult with the real-world data WIMOB and the NABM designed in chapter 1 Figure 2.2. Since we do not have an access to individual-based reports, we can only measure the performance based on simulated data attained from running the calibrated model.

Let Ψ denote the NABM model, let ξ denote the optimal set of parameters after calibrating the model to the surveillance data in any campus defined in Chapter 1. The return of Ψ is a forest indicating the ground truth infection cascade in the same schema of the return of Cult. We use perturb(Y, γ) to simulate the real-world individual-based report concerning the rate of false negative for any disease testing. perturb(Y, γ) will randomly remove γ % of the nodes in the forest Y. The pseudo code for validating Cult with our data is presented in Algorithm 3:

The Metric(Y_1, Y_2) function is for measuring the similarity between two sets Y_1 and Y_2. For instance, we can use Jaccard coefficient over the nodes covered by the two forests [79].

Algorithm 3 Cult-validation

procedure CULT($G(V, E), F, C, D, \alpha, \Psi, \xi, \gamma$)
    $M = \phi$
    for $i = 1, 2, 3...W$ **do**
        $X_i = \Psi(\xi)$
        $S_i = \text{perturb}(X_i, \gamma)$
        $T_i = \text{Cult}(G(V, E), F, F, S_i, \alpha)$
        $M \cup \text{Metric}(X_i, T_i)$
    return M

At this time, we have not completed the whole validation for Cult-sampling yet, we will present the entire setup of validation for Cult-sampling in our future publications.

4.7 A Case Study: GT

In this section, we will implement Algorithm 2 over the group-based testing data for Covid-19 given to us at GT campus in Fall 2020 semester. In particular, each group in this case is defined as a collection of students living in the same residence or fraternity building.

4.7.1 Residential Groups' Report

In the fall semester of 2020, the team conducting the surveillance testing program at GT was working closely with the campus officials to monitor an outbreak of Covid-19 that appeared among residential students [49]. For simplification, we select a week when surveillance testing program finally identified 25 cases among all residential groups, including fraternity houses, from 10/24/2020 to 10/30/2020 on campus. The report shows the number of positive cases for each residence building or fraternity hall every day.

4.7.2 Residential Individuals

Since the original WIMOB data is anonymized, we can only infer residential individuals from the network based on a designed rule. We first explore all active users among the bipartite network from the move-in week of all residential students, from 08/14/2020 to

08/20/2020. Note that we define residential groups based on the IDs of all residence halls and fraternity houses. Second, we set up the rule as follows: anyone staying in any residence hall or fraternity house with the most prolonged duration greater than 60 minutes from 6 pm to 10 am of the next day of any working day is an inferred resident within the residential group corresponding to this building. In particular, we define the working days here as a week without the weekend and Friday. Our analysis finally identifies about 6800 residential students, and among them, about 1200 living in fraternity houses in the Fall 2020 semester.

4.7.3 Setup

We have selected the input D and $J_1, J_2...J_n$ based on our data. We are also given the collocation network $G = (V, E)$ from 08/14/2020 to 08/20/2020. Moreover, to start with a source size $k = 3$ which is the total reported cases in 08/13/2020, we set $\alpha = 50$. To respect on individual's privacy, instead of showing the individual's infection cascade as outputted by Algorithm 2, we only visualize the most frequent (top 15) cross-groups transmissions from the forests $T_1, T_2, ..., T_W$.

4.7.4 Result

Our experiment finally identifies 80 ± 7 cases. Among them, 25 cases have been reported by the data given to us. About 70% of the nodes in each infection cascade are missed by the surveillance testing at GT.

Note that all residence and fraternity buildings are distributed exactly on the geographical locations based on the campus map in the following plots. Figure 4.1 shows that the disease transmits between buildings at the lower right corner on campus on 10/24/2020. Then, on the second day of the week, it reached the upper left corner of the campus Figure 4.2. On 10/27/2020, all the building in the upper left corner had been source of transmissions Figure 4.4. On 10/28/2020 and 10/29/2020, Friday and Saturday, the disease was brought

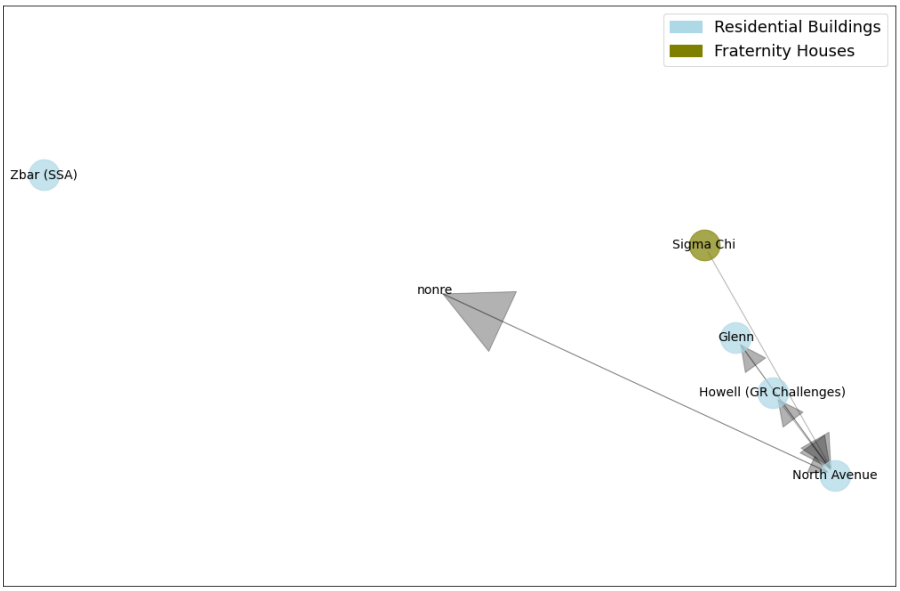


Figure 4.1: 10/24/2020, ”nonre” is the group for all non-residential individuals

into fraternity houses from other residential buildings on campus Figure 4.5, Figure 4.6.

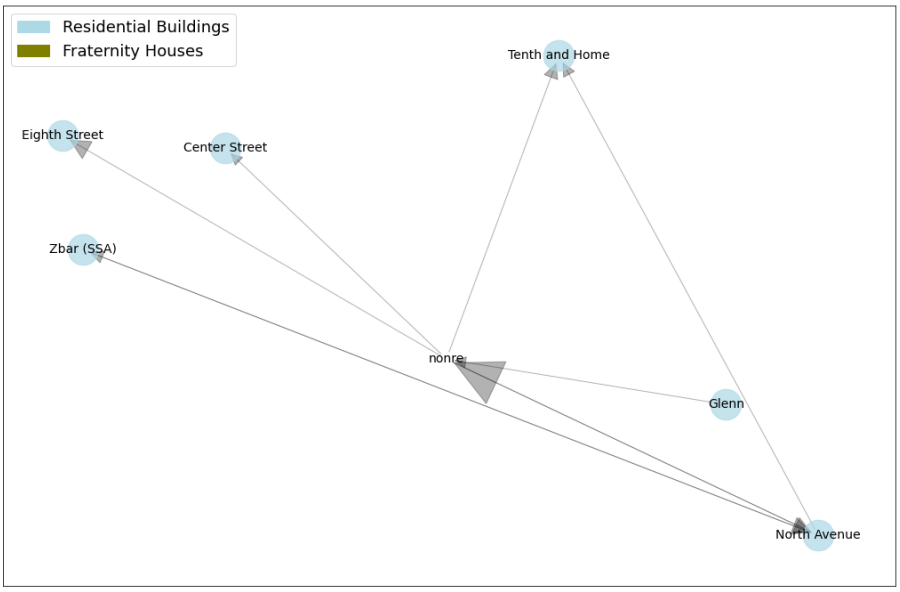


Figure 4.2: 10/25/2020

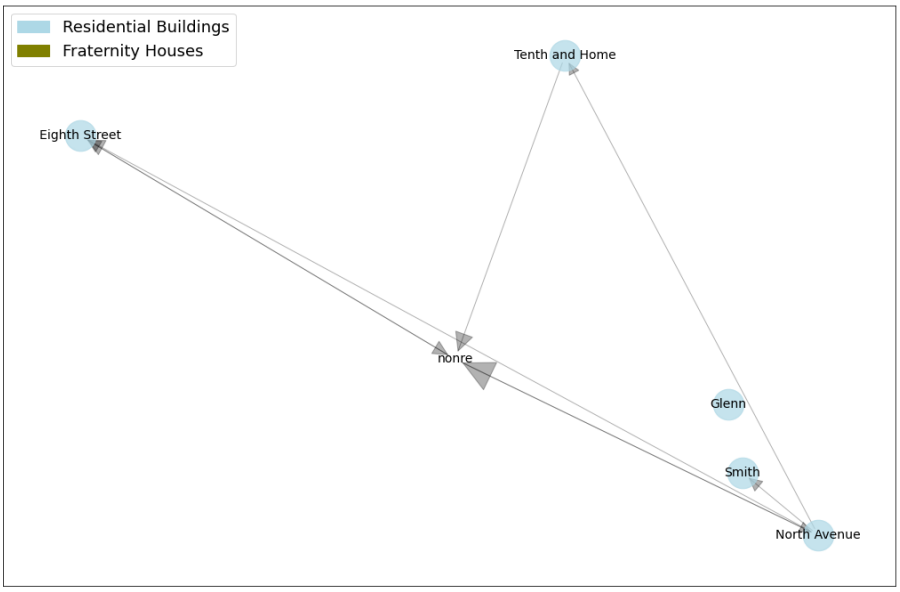


Figure 4.3: 10/26/2020

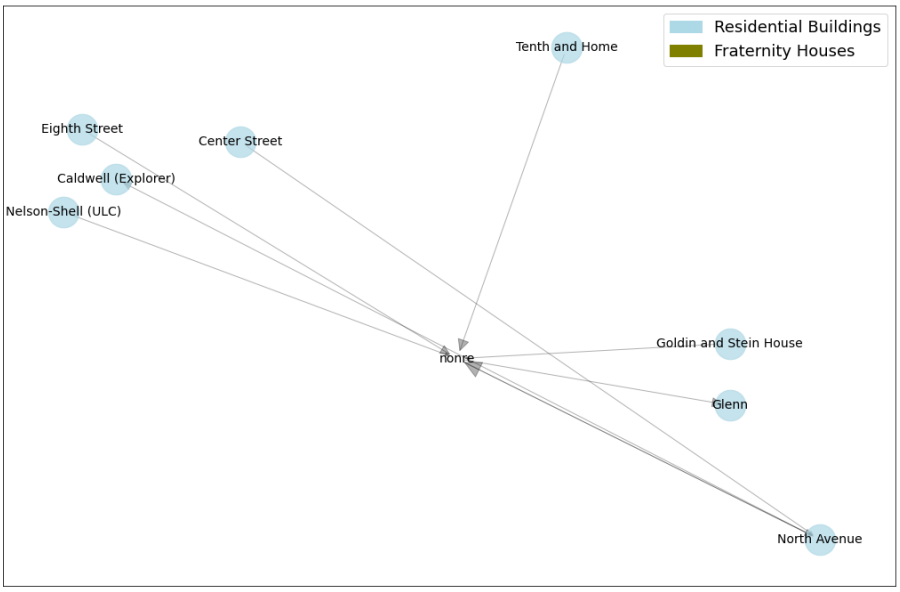


Figure 4.4: 10/27/2020

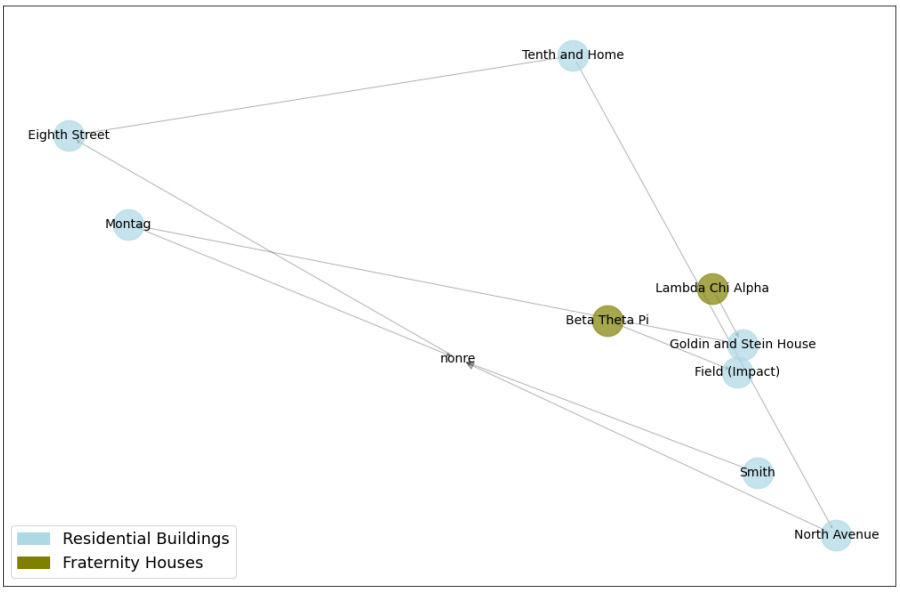


Figure 4.5: 10/28/2020

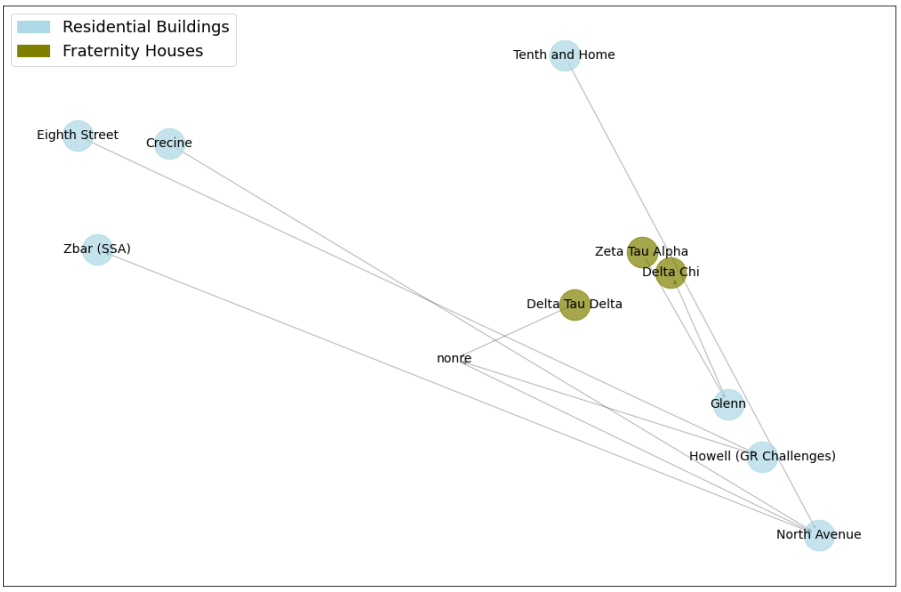


Figure 4.6: 10/29/2020

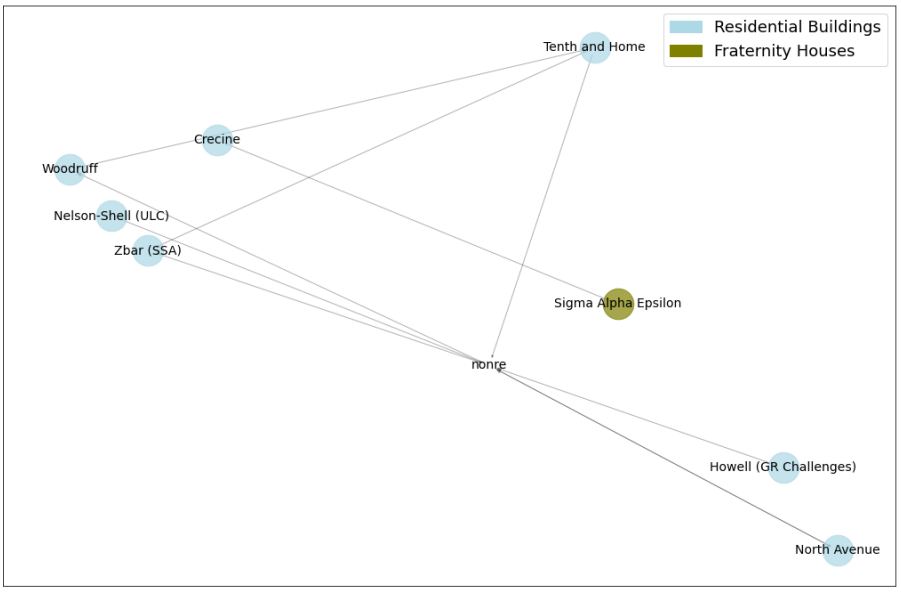


Figure 4.7: 10/30/2020

CHAPTER 5
CONCLUSION, DISCUSSION, AND FUTURE WORK

5.1 Conclusion

In the long run, universities worldwide need to find a way to operate safely during an infectious disease epidemic. For instance, when the Covid-19 emerged on campus, universities sought to design and improve nonpharmaceutical interventions (NPI) strategies to minimize the size of a potential outbreak. In addition, they might want to know what the chains of transmissions and sources are causing the current infection. Answer these questions typically requires disease modeling since models help abstract the disease dynamics and reason more about the mechanism of disease transmission among campus communities. However, most current modeling works on university campuses rely on aggregated data, such as Course enrollment data (EN) or testing statistics [17, 10, 11, 12]. However, EN is static in general, with a limited scope of information given from registration data only among students and updated non-frequently. Therefore, EN cannot describe the real-world population dynamics on campus. Besides, surveillance testing on campus can be noisy due to sampling bias and false-negative rates [49]. The policies yielded by the disease models using only the aggregated data are considered both inaccurate and burdensome to both individuals and universities.

We propose using the on-campus managed WiFi to proximate human mobility for disease modeling and policy design and evaluation. We demonstrate that WiFi Mobility (WIMOB) can assist universities in designing and evaluating strategies for controlling infections on campus. In particular, using WIMOB to approximates the contact network among individuals, we construct and validate a network-based simulation model of Covid-19 on a university campus. Then, we design and evaluate two novel methods for improving

decision-making powered by the WiFi mobility data with the model constructed. The first method outputs a more granular and localized closure policy, which is more effective in terms of disease intervention outcomes and less burdensome to individuals and schools. The second method can reliably reconstruct likely chains of transmissions over time, given the current noisy testing statics.

5.1.1 Policy Design and Evaluation

NPI are the first line of defense for universities to respond to contagious diseases like COVID-19 [80, 81], and are also crucial to control infections and continue operations till recovery. On a campus, a common form of NPI is closure [82]. Universities consider EN to design remote instruction (remote instruction policies (RI)) for closure to support continued operations safely [17]. However, EN can misconstrue contact on campus and RI policies can have broad impacts despite their effects on curbing the disease spread. This chapter demonstrates that repurposing logs from a managed WiFi network (WIMOB) can help design effective localized closure policies (Localized closure policy (LC)). We simulate COVID-19 with an network agent-based Models (NABM) that harnesses WIMOB to compare RI and LC. As universities plan for Fall 2021, our results present evidence that LC can lead to improved infection reduction outcomes, while simultaneously relaxing burdens on the campus caused by coarse-grained broad RI policies.

5.1.2 Reconstruct An Epidemic over Time

Universities are interested in reconstructing the epidemic by discovering the sources and flows causing the outbreak among individuals on campuses, given noisy testing data. Existing studies have already presented various formulations and corresponding efficient algorithms which guarantee good approximation to the optimal solution [72, 74, 71, 70]. However, we realize that these formulations are far from reality considering the input data given to us. We propose a formulation Group-level-influence that leverages existing work

[75]. The inputs are similar to the real-world data available: (1) a time-varying collocation network and (2) a noised surveillance data reporting the number of positive cases for groups. In particular, we introduce Algorithm 2, a sampling procedure to convert the group's data into individual-based reports to solve our problem by calling an existing algorithm multiple times. We also design a process to validate Algorithm 1 with our WIMOB and NABM. Lastly, we present a case study over the group-based data of Covid-19 on residential groups at Georgia Institute of Technology (GT) in Fall 2020 semester.

5.2 Discussion

5.2.1 WIMOB Data

Operational Considerations: Beyond university campuses, any public institutes, such as airports, companies, or hospitals, can devise practical methods of obtaining, storing, and processing mobility of the community as WIMOB. Information teams can access logs from the managed network internally as it is passively collected. Moreover, it does not require any new form of surveillance sensing, but administrators must revise terms of use and stay sensitive to community perspectives. Despite population mobility being valuable for many applications [83], accumulating localization data can be a significant privacy concern [84]. Instead, operational applications need to conceive approaches that only retain insights on locations to shutdown but not individual data. Similarly, any practical use needs to have pre-established access limitations on what stakeholders can learn from the data [85] (e.g., decision-makers can only get a list of candidate locations to close).

Privacy, Ethics and Legal Considerations: We purposefully compare our prototype targeted policies against moving classes online because of practical budgets within the university. Both the WIMOB and EN based contact networks are derived from archival data accumulated by universities. This does not require instrumenting campus or its community with any new form of surveillance infrastructure. However, its use for a different purpose demands approval by an IRB. Moreover, acquiring these kinds of data would require col-

laborating with data-stewards (e.g., the IT department) to establish a data-use agreement. This document must clarify how the data will be de-identified, transferred, and stored.

For this form of data, the critical privacy challenge might not be localization itself, but rather the aggregation of data over a period of time [84]. Data spanning a longer period are more susceptible to cross-analyzing and identifying. To mitigate over-accumulation of data, we suggest an adherence to principles of data minimization [86]. Instead of storing entire mobility graphs, the campus can compute and preserve only high-level insights, such as the importance of locations. This redacts any underlying individual behavior and corresponding identifiable information. Actually, for future purposes campuses can consider a form of differential privacy that authorizes limited forms of data querying depending on the privileges of the stakeholder [85].

An operational application would require the university to update the terms of use for its managed network. Particularly, the university should disclose how this data can be used in critical circumstances that invoke shared vulnerabilities [87]. On notifying the campus community of this change it offers individuals the choice to refrain from using the university network. Prior work on a sample within the same university campus shows that 90% of students are connected to the network on any given day [38]. Therefore, proposing such an opt-out condition can be viewed as an unfair choice. As a result, the campus needs to develop a contingency plan to accommodate network access to users who do not want their mobility behavior to constitute the aggregated insights.

5.2.2 Policy Making

Recommendations for Campus Administrators (i) Policymakers should pay attention to academic and student services spaces under an epidemic. Recall that in Chapter 3, we have illustrated that LC shuts down fewer locations on campus and identifies a wider variety of auxiliary spaces, see Figure A.17 and Figure A.18. An observation is that WiMob not only target sites at the level of rooms, suites, or labs but also give insights into the closed

locations. In all scenarios and budgets, both RI and LC mostly shut down academic and student-service spaces. Irrespective of the percentage of these areas, other dimensionalities might also explain why shutting down these locations. For example, the topology of rooms, air condition, and schedules might contribute to novel epidemiological findings on campus. Although this thesis proposes using WIMOB to inform LC, it is flexible for policymakers to choose any data at the lower granularity but with diverse location types on campuses.

(ii) Contact-tracing teams on campus should use technologies to identify chains of transmission. Current contact-tracing spends a great effort identifying past contacts by survey, given a positive case, and assign test kits to these individuals [88, 4]. In Chapter 4, we have seen Algorithm 1 can be applied to discover transmission lines and missing infections from the testing reports. If individual-based reports are given, the contact-tracing team can leverage this technique to identify likely missing infections quickly to prioritize them to the testing service.

Generalizability of Closure Policies for Other Contexts: In practice LC policies should be deployed on campus in conjunction with the other tools as well like testing, tracing, and quarantining. WIMOB can complement disease-specific knowledge to identify closure spaces. For example, small indoor spaces with poor ventilation increase the risk of infection for COVID-19 [89], while other algorithm-identified locations for closure might not require closure because mask-wearing and testing have high compliance among users of that space. Further, as a pandemic progresses and public health guidance develops [90], with WIMOB, campuses can regulate the restriction of LC policies and anticipate the path to 'normal' operations [91, 92]. Moreover, WIMOB captures various spillover effects that cannot be captured in methods like EN. For instance, with WIMOB we observe that the mobility in Fall 2020 was 39% of that in Fall 2019 because the on-ground policies lead to certain staff working remotely as well. With additional information, WIMOB enables policymakers to model such scenarios and design alternatives like LC with new budgets. Policymakers can use WIMOB as a versatile tool to explore dynamic intervention

strategies as well. Prior work shows that staggering policy restrictions could have variable impact on campus [93]. Accordingly, WIMOB could be used to build an adaptive version of LC that updates at different points in the semester based on expected mobility changes. Additionally, depending on campus priorities and resource limitations, different campuses can use this same data to model policies differently. The effectiveness of reopening policies is expected to be sensitive to a campus' specific context that includes physical infrastructure, overarching guidelines, and human compliance [94]. For certain campuses policies might not need to be constrained by exposure risk as testing might be frequent, ubiquitous, and voluminous. Other campuses could have limits on quarantining capacity. Policymakers might even consider the cost trade-offs by actually forecasting actual financial losses incurred by reduction in mobility [95], or valuate loss of services based on community needs [96].

5.3 Limitation and Future Work

This work presents evidence that university campuses can repurpose existing data sources to inform the design of policies that can control COVID-19. One of the drawbacks of WIMOB, however, is that it assumes all edges to be the same. For example, when constraining by mobility, in real scenarios losing certain visits might be more valuable than others. Decline in mobility around profit-making services, such as shops and cafeterias, versus losing mobility at common rooms have a different tangible effects on campus. Currently, we take an agnostic stance towards the mobility behavior, where all visits at all locations are the same. In reality, implementing policies could have inequitable qualitative impacts despite appearing to have a similar network configuration. This can be improved by embedding more qualitative information into the network and conceiving ingenious ways to associate costs to edges.

Similar to the assumption that all visits and locations, the current work also assumes all people to be equal. However, different people have different underlying conditions

that can make their vulnerabilities more concerning [69]. The privacy safeguards of this study restricted the research team from acquiring any additional demographic or historical information. Further work can attempt to characterize the nodes by randomly seeding the network to reflect the approximate demographic break up of the community. Alternatively, researchers could try to estimate some demographic based on behavior as well. However, to leverage accurate individual information, even for operational use during a public health emergency, policymakers and researchers need to develop new privacy protocols [97].

Also, Chapter 3 only studies three rudimentary scenarios, *persistence*, *non-residential avoidance*, *complete avoidance*. These scenarios assume that when a location is shutdown, the individuals who ought to have visited that location do not come into contact with anyone else during the same time. Yet, other substitution behaviors are possible and the richness of networks leveraged with WIMOB enables the exploration of various new scenarios that can be triggered by policy interventions on campus. For instance, individuals might not even visit transitory spaces, such as lobbies or cafes between classes. Certain collocations could be the consequence of social ties which might never be developed because of a shutdown (e.g., project teams meeting outside of class). Further research can illuminate the effects of policies in more specific scenarios by modeling post-intervention behavior more accurately.

Lastly, our work for the validation of Algorithm 2 remains incomplete. Further work should focus on how to aggregate the trees from iteratively running Algorithm 1 with sampled individual reports.

Appendices

APPENDIX A

SUPPLEMENTARY RESULTS OF CHAPTER 3

Table A.1: Calibration outcomes with variations. The results in the main paper use variables p, α, and I_0 as estimated by calibrating the simulation model on the first 5 weeks of positivity rates provided by GT surveillance for Fall 2020, while incorporating external cases from Fulton County. For sensitivity analyses, we perform calibrations on GT data for weeks $5-9$ and $10-14$. Additionally, we perform calibrations on first five weeks of UIUC and Berkeley positivity rate (along with data from their respective county). These parameters were found by validating the NABM on the remaining weeks of Fall 2020. To assess the basic reproductive number (R_0) of our NABM we study the first 4 weeks of the disease. We find the effective R_0 to be higher for Fall 2019 than Fall 2020 as the mobility behaviors between the 2 semesters was vastly different. Note, Fall 2020 exhibits only 39% of the mobility we observe in Fall 2019. In fact, the NABM is calibrated on Fall 2020, where behavior was subject to pandemic related closures, but in Fall 2019 the mobility was not hindered by any interventions. Thus, Fall 2019 reflects a counterfactual of Fall 2020 without any closures.

	Calibrating on Positivity Rate at GT			Calibrating on Positivity Rate with other University Behavior	
Parameter	weeks $0-4$	weeks $5-9$	weeks $10-14$	UIUC	Berkeley
p	0.034 ± 0.007	0.073 ± 0.005	0.0024 ± 0.0003	0.024 ± 0.0009	0.041 ± 0.003
α	0.032 ± 0.0032	0.0042 ± 0.0006	0.0159 ± 0.002	0.0069 ± 0.0013	0.038 ± 0.006
I_0	0.012 ± 0.0009	0.00057 ± 0.00007	0.0030 ± 0.0007	0.0039 ± 0.0013	0.0048 ± 0.0003
Optimal r.m.s.e	0.0034	0.0007	0.0015	0.0028	0.0031
Effective R_0 (min - max), Fall 2020	$1.15-1.18$	$1.17-2.14$	$0.33-0.95$	$1.12-1.19$	$1.24-1.28$
Effective R_0 (min - max), Fall 2019	$2.87-5.68$	$5.15-12.93$	$1.27-1.36$	$3.35-5.35$	$3.32-7.00$

Table A.2: Comparison of different Localized closure policies with Pagerank ($\mathsf{LC}_{\text{PRank}}$) policies in terms of controlling the disease and impacts on campus in Fall 2019; calibrated from week $0-4$ in Fall 2020 at GT. Note that this table is the same as Table 3.1. We repeat the results here for easier comparison of $\mathsf{LC}_{\text{PRank}}$ to other algorithms shown in Table A.3, Table A.4 and Table A.5. Within each scenario, we perform the Kruskal-Wallis H-Test [66] to compare outcomes of $\mathsf{LC}_{\text{PRank}}$ with RI. We find that $\mathsf{LC}_{\text{PRank}}$ leads to significantly improved peak infection reduction and internal transmission. In terms of reduction in total infections, the outcomes are comparable in general but can vary by specific scenarios. In addition, every policy also exerts some burden on campus, either in terms of locations affected, students avoiding campus or isolation. We observe that $\mathsf{LC}_{\text{PRank}}$ policies focus on fewer locations (except in Complete Avoidance (*S3*)). Moreover, these policies affect fewer student's schedules and therefore fewer people avoid campus due to completely remote schedules. Finally, $\mathsf{LC}_{\text{PRank}}$ does not increase the percentage of people completely isolated on campus (p-value: < 0.01:*, < 0.001:**).

Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	RI	LC		RI	LC		RI	LC	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	25.34(±12)	36.92(±14)**	34.30(±13)**	35.44(±10)	49.33(±11)**	52.19(±10)**	61.62(±7)	69.34(±5)**	64.44(±6)**
Total Infections (%)	6.99(±5)	10.63(±6)**	8.19(±5)**	14.88(±4)	13.96(±6)*	15.67(±6)	33.00(±5)	33.4(±5)	26.94(±5)**
Internal Transmissions (%)	17.13(±9)	22.62(±11)**	21.01(±11)**	27.58(±8)	35.35(±12)**	39.20(±11)**	54.00(±8)	70.89(±7)**	60.90(±9)**
Burdens on Campus									
Locations Affected	58	18	19	58	38	50	58	192	124
Students Avoiding (%)	0	0	0	9.30	0.20	0.45	27.21	12.45	6.57
Completely Isolated on Campus (%)	5.42	8.40	8.40	5.95	5.72	5.71	7.09	5.18	5.23

Table A.3: Comparison of different Localized closure policies with Betweenness Centrality ($\mathsf{LC}_{\text{BCen}}$) policies in terms of controlling the disease and impacts on campus in Fall 2019; calibrated from week $0-4$ in Fall 2020 at GT. Within each scenario, we perform the Kruskal-Wallis H-Test [66] to compare outcomes of $\mathsf{LC}_{\text{BCen}}$ with RI. We find that $\mathsf{LC}_{\text{BCen}}$ leads to significantly improved peak infection reduction and internal transmission, when designed with the exposure risk budget, but can be worse with the mobility budget. In terms of reduction in total infections, the outcomes are typically worse. In addition, every policy also exerts some burden on campus, either in terms of locations affected, students avoiding campus or isolation. We observe that $\mathsf{LC}_{\text{BCen}}$ policies focus on fewer locations (except in *S3*). Moreover, these policies affect fewer student's schedules and therefore fewer people avoid campus due to completely remote schedules. Finally, Localized closure policies with Load Centrality ($\mathsf{LC}_{\text{LCen}}$) does not increase the percentage of people completely isolated on campus (p-value: < 0.01:*, < 0.001:**).

Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	RI	$\mathsf{LC}_{\text{BCen}}$		RI	$\mathsf{LC}_{\text{BCen}}$		RI	$\mathsf{LC}_{\text{BCen}}$	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	25.34(±12)	19.14(±12)**	30.93(±13)**	35.44(±10)	30.79(±13)**	51.87(±10)**	61.62(±7)	65.07(±6)**	61.38(±7)
Total Infections (%)	6.99(±5)	4.85(±4)**	7.74(±5)	14.88(±4)	7.76(±5)**	15.30(±6)	33.00(±5)	25.32(±5)**	22.08(±6)**
Internal Transmissions (%)	17.13(±9)	11.96(±9)**	19.64(±10)**	27.58(±8)	19.63(±10)**	38.74(±11)**	54.00(±8)	63.29(±8)**	54.00(±8)
Burdens on Campus									
Locations Affected	58	18	19	58	38	50	58	192	124
Students Avoiding (%)	0	0	0	9.30	0.07	0.45	27.21	11.47	6.74
Completely Isolated on Campus (%)	5.42	8.63	8.63	5.95	5.49	5.47	7.09	5.15	5.19

Table A.4: Comparison of different Localized closure policies with Eigenvector Centrality ($\mathsf{LC}_{\text{ECen}}$) policies in terms of controlling the disease and impacts on campus in Fall 2019; calibrated from week $0-4$ in Fall 2020 at GT. Within each scenario, we perform the Kruskal-Wallis H-Test [66] to compare outcomes of $\mathsf{LC}_{\text{ECen}}$ with RI. We find that $\mathsf{LC}_{\text{ECen}}$ leads to significantly improved peak infection reduction and internal transmission. In terms of reduction in total infections, the outcomes vary by specific scenarios. In addition, every policy also exerts some burden on campus, either in terms of locations affected, students avoiding campus or isolation. We observe that $\mathsf{LC}_{\text{ECen}}$ policies focus on fewer locations (except in *S3*). Moreover, these policies affect fewer student's schedules and therefore fewer people avoid campus due to completely remote schedules. Finally, $\mathsf{LC}_{\text{ECen}}$ does not increase the percentage of people completely isolated on campus (p-value: < 0.01:*, < 0.001:**).

Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	RI	$\mathsf{LC}_{\text{ECen}}$		RI	$\mathsf{LC}_{\text{ECen}}$		RI	$\mathsf{LC}_{\text{ECen}}$	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	25.34($\pm$12)	36.15($\pm$13)**	36.13($\pm$13)**	35.44($\pm$10)	44.52($\pm$12)**	51.33($\pm$10)**	61.62($\pm$7)	65.13($\pm$6)**	62.15($\pm$7)
Total Infections (%)	6.99($\pm$5)	8.66($\pm$6)**	8.69($\pm$6)**	14.88($\pm$4)	11.75($\pm$6)**	14.96($\pm$6)	33.00($\pm$5)	25.39($\pm$5)**	22.82($\pm$6)**
Internal Transmissions (%)	17.13($\pm$9)	22.33($\pm$11)**	22.37($\pm$11)**	27.58($\pm$8)	29.95($\pm$12)*	37.94($\pm$11)**	54.00($\pm$8)	63.56($\pm$8)**	57.07($\pm$10)**
Burdens on Campus									
Locations Affected	58	18	19	58	38	50	58	192	124
Students Avoiding (%)	0	0	0	9.30	0.20	0.55	27.21	13.11	6.96
Completely Isolated on Campus (%)	5.42	8.59	8.59	5.95	5.53	5.51	7.09	5.17	5.23

Table A.5: Comparison of different $\mathsf{LC}_{\text{LCen}}$ policies in terms of controlling the disease and impacts on campus in Fall 2019; calibrated from week $0-4$ in Fall 2020 at GT. Within each scenario, we perform the Kruskal-Wallis H-Test [66] to compare outcomes of $\mathsf{LC}_{\text{LCen}}$with RI. We find that $\mathsf{LC}_{\text{LCen}}$leads to significantly improved peak infection reduction and internal transmission. In terms of reduction in total infections, the outcomes are comparable in some scenarios but can vary in specific scenarios. In addition, every policy also exerts some burden on campus, either in terms of locations affected, students avoiding campus or isolation. We observe that $\mathsf{LC}_{\text{LCen}}$policies focus on fewer locations (except in *S3*). Moreover, these policies affect fewer student's schedules and therefore fewer people avoid campus due to completely remote schedules. Finally, $\mathsf{LC}_{\text{LCen}}$does not increase the percentage of people completely isolated on campus (p-value: < 0.01:*, < 0.001:**).

Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	RI	$\mathsf{LC}_{\text{LCen}}$		RI	$\mathsf{LC}_{\text{LCen}}$		RI	$\mathsf{LC}_{\text{LCen}}$	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	25.34(±12)	22.42(±13)**	30.73(±13)**	35.44(±10)	32.85(±13)*	51.44(±10)**	61.62(±7)	65.01(±6)**	61.40(±7)
Total Infections (%)	6.99(±5)	5.48(±5)**	7.64(±5)	14.88(±4)	8.23(±5)**	15.03(±6)	33.00(±5)	25.33(±5)**	21.98(±6)**
Internal Transmissions (%)	17.13(±9)	13.79(±9)**	19.37(±10)**	27.58(±8)	20.86(±11)**	38.08(±11)**	54.00(±8)	63.28(±8)**	55.28(±9)
Burdens on Campus									
Locations Affected	58	18	19	58	38	50	58	192	124
Students Avoiding (%)	0	0	0	9.30	0.07	0.43	27.21	11.47	6.73
Completely Isolated on Campus (%)	5.42	8.63	8.63	5.95	5.49	5.47	7.09	5.15	5.20

Table A.6: Comparison of different $\mathsf{LC}_{\text{PRank}}$ policies in terms of controlling the disease and impacts on campus in Fall 2019; calibrated from week $5-9$ in Fall 2020 at GT. Within each scenario, we perform the Kruskal-Wallis H-Test [66] to compare outcomes of $\mathsf{LC}_{\text{PRank}}$ with RI. We find that $\mathsf{LC}_{\text{PRank}}$ leads to significantly improved peak infection reduction and internal transmission. In terms of reduction in total infections, the outcomes are better in general but can be comparable in specific scenarios. The burden exerted on campus is the same as structural impacts of $\mathsf{LC}_{\text{PRank}}$ (Table A.2). (p-value: < 0.01:*, < 0.001:**).

Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	Broad	$\mathsf{LC}_{\text{PRank}}$		RI	$\mathsf{LC}_{\text{PRank}}$		RI	$\mathsf{LC}_{\text{PRank}}$	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	$20.10(\pm4)$	$25.60(\pm3)^{**}$	$25.63(\pm3)^{**}$	$31.25(\pm3)$	$42.32(\pm4)^{**}$	$47.29(\pm4)^{**}$	$62.35(\pm2)$	$88.87(\pm2)^{**}$	$76.89(\pm3)^{**}$
Total Infections (%)	$8.89(\pm2)$	$10.50(\pm3)^{**}$	$9.70(\pm3)^{**}$	$20.26(\pm2)$	$20.02(\pm3)$	$23.71(\pm4)^{**}$	$46.72(\pm2)$	$67.92(\pm4)^{**}$	$51.30(\pm4)^{**}$
Internal Transmissions (%)	$9.97(\pm2)$	$11.51(\pm2)^{**}$	$10.95(\pm2)^{**}$	$21.84(\pm2)$	$22.51(\pm3)$	$26.64(\pm3)^{**}$	$49.80(\pm2)$	$74.96(\pm3)^{**}$	$56.89(\pm4)^{**}$

Table A.7: Comparison of different $\mathsf{LC}_{\text{PRank}}$ policies in terms of controlling the disease and impacts on campus in Fall 2019; calibrated from week $10-14$ in Fall 2020 at GT. Within each scenario, we perform the Kruskal-Wallis H-Test [66] to compare outcomes of $\mathsf{LC}_{\text{PRank}}$ with RI. We find that $\mathsf{LC}_{\text{PRank}}$ leads to significantly improved peak infection reduction and internal transmission. In terms of reduction in total infections, the outcomes are better in general but can be comparable in specific scenarios. The burden exerted on campus is the same as structural impacts of $\mathsf{LC}_{\text{PRank}}$ (Table A.2). (p-value: < 0.01:*, < 0.001:**).

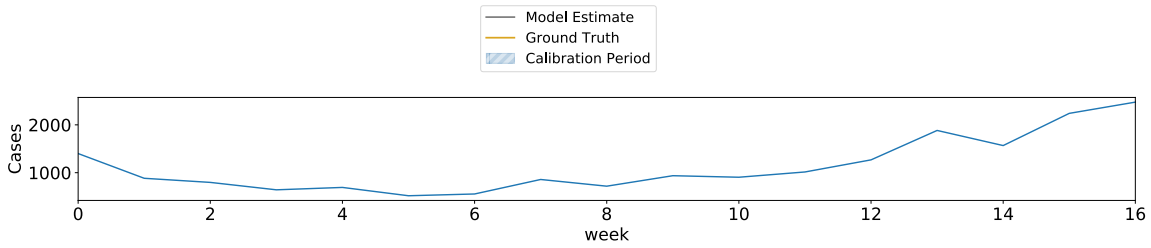
Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	Broad	$\mathsf{LC}_{\text{PRank}}$		RI	$\mathsf{LC}_{\text{PRank}}$		RI	$\mathsf{LC}_{\text{PRank}}$	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	$-1.75(\pm 8)$	$3.65(\pm 8)^{**}$	$-1.95(\pm 8)$	$3.88(\pm 8)$	$-2.24(\pm 8)^{**}$	$-2.06(\pm 8)^{**}$	$20.39(\pm 7)$	$7.57(\pm 8)^{**}$	$2.81(\pm 8)^{**}$
Total Infections (%)	$3.93(\pm 9)$	$10.36(\pm 8)^{**}$	$5.13(\pm 9)$	$9.87(\pm 8)$	$6.36(\pm 9)^{**}$	$6.48(\pm 9)^{**}$	$26.02(\pm 7)$	$16.37(\pm 8)^{**}$	$11.80(\pm 8)^{**}$
Internal Transmissions (%)	$42.33(\pm 10)$	$61.15(\pm 7)^{**}$	$56.25(\pm 8)^{**}$	$49.83(\pm 9)$	$67.10(\pm 6)^{**}$	$69.10(\pm 6)^{**}$	$74.74(\pm 5)$	$84.80(\pm 3)^{**}$	$79.90(\pm 4)^{**}$

Table A.8: Comparison of different $\mathsf{LC}_{\text{PRank}}$ policies in terms of controlling the disease and impacts on campus in Fall 2019; calibrated from week $0-4$ in Fall 2020 at UIUC. Within each scenario, we perform the Kruskal-Wallis H-Test [66] to compare outcomes of $\mathsf{LC}_{\text{PRank}}$ with RI. We find that $\mathsf{LC}_{\text{PRank}}$ leads to significantly improved peak infection reduction, internal transmission and total infections. The burden exerted on campus is the same as structural impacts of $\mathsf{LC}_{\text{PRank}}$ (Table A.2). (p-value: < 0.01:*, < 0.001:**).

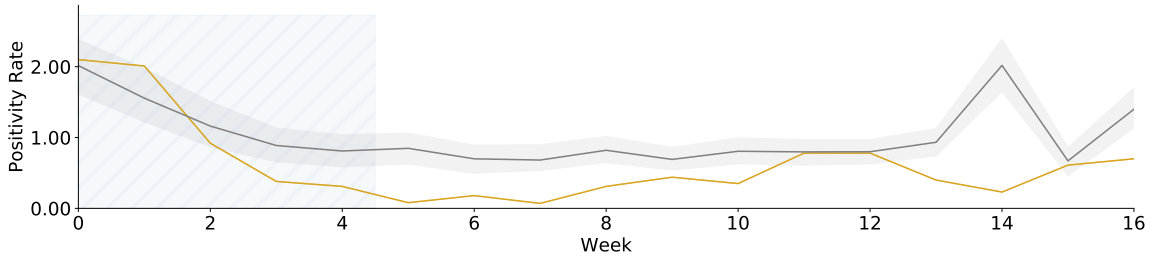
Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	Broad	$\mathsf{LC}_{\text{PRank}}$		RI	$\mathsf{LC}_{\text{PRank}}$		RI	$\mathsf{LC}_{\text{PRank}}$	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	41.40(±3)	60.44(±2)**	59.52(±2)**	49.75(±2)	74.22(±2)**	76.44(±2)**	78.14(±1)	85.81(±1)**	83.71(±1)**
Total Infections (%)	18.46(±3)	27.12(±3)**	25.25(±3)**	27.09(±3)	38.00(±4)**	40.68(±4)**	51.97(±3)	59.93(±5)**	54.07(±5)**
Internal Transmissions (%)	28.22(±3)	40.93(±3)**	39.09(±3)**	37.89(±3)	58.47(±2)**	65.45(±2)**	68.04(±2)	86.45(±1)**	80.08(±1)**

Table A.9: Comparison of different $\mathsf{LC}_{\text{PRank}}$ policies in terms of controlling the disease and impacts on campus in Fall 2019; calibrated from week $0-4$ in Fall 2020 at UC Berkeley. Within each scenario, we perform the Kruskal-Wallis H-Test [66] to compare outcomes of $\mathsf{LC}_{\text{PRank}}$ with RI. We find that $\mathsf{LC}_{\text{PRank}}$ leads to significantly improved peak infection reduction, internal transmission and total infections. The burden exerted on campus is the same as structural impacts of $\mathsf{LC}_{\text{PRank}}$ (Table A.2). (p-value: < 0.01:*, < 0.001:**).

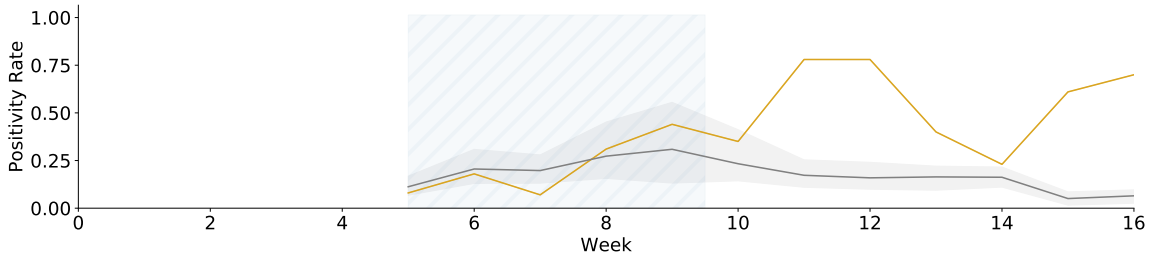
Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	Broad	$\mathsf{LC}_{\text{PRank}}$		RI	$\mathsf{LC}_{\text{PRank}}$		RI	$\mathsf{LC}_{\text{PRank}}$	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	29.13(±3)	36.46(±5)**	36.34(±5)**	38.83(±3)	54.95(±4)**	58.88(±4)**	66.69(±2)	78.18(±1)**	77.65(±2)**
Total Infections (%)	6.34(±3)	8.59(±3)**	7.28(±3)**	14.71(±3)	13.18(±4)**	14.83(±4)	33.86(±4)	33.98(±5)	27.10(±5)**
Internal Transmissions (%)	15.99(±3)	20.43(±4)**	19.17(±4)**	27.01(±3)	34.60(±4)**	38.78(±4)**	55.01(±2)	74.65(±2)**	63.57(±3)**



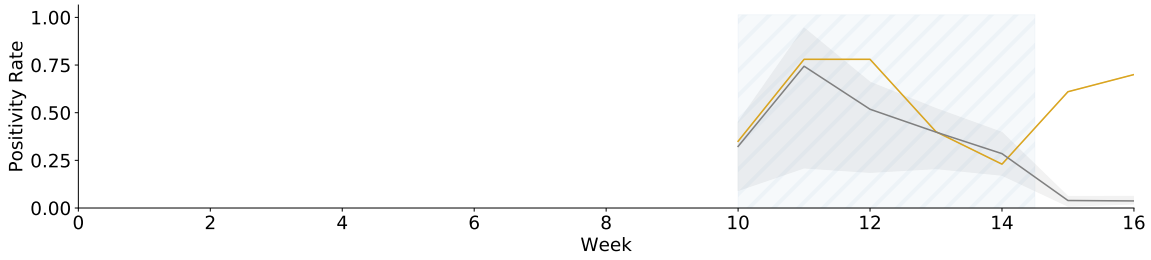
(a) External cases



(b) Calibrating on the weeks 0-4

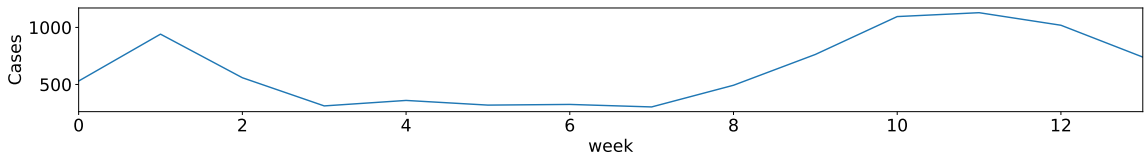


(c) Calibrating on the weeks 5-9

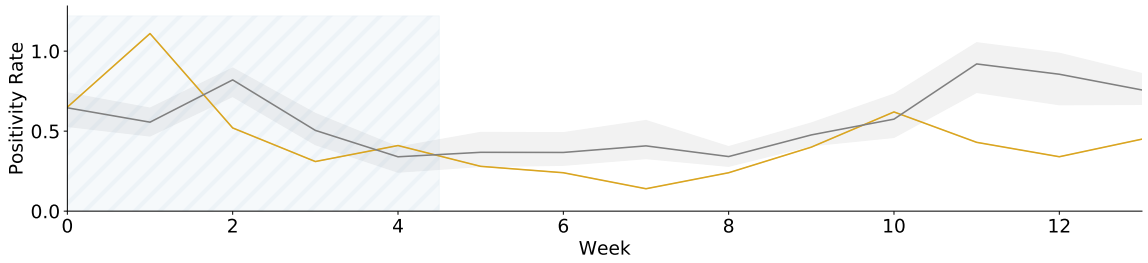


(d) Calibrating on the weeks 10-14

Figure A.1: We calibrate NABM on positivity rates from Fall 2020 at GT. The objective function of the calibration is to minimize the r.m.s.e. with the weeekly average of positivity rate obtained from surveillance testing results at GT [49]. (a) The parameter that determines external transmission of infections on a given day, $I_{out}(t)$, is a function of cases in Fulton county (where GT is located). (b) The models discussed in the main paper are calibrated using the first 5 weeks of data. We illustrate the output for a range of parameters that incorporate quantitative uncertainty, i.e., within 40% of the r.m.s.e. (c, d) illustrate calibration on the second period of 5 weeks and third period of 5 weeks respectively. These only show the optimal parameter output. The shaded region around the lines show the 2.5^{th} and 97.5^{th} percentile.

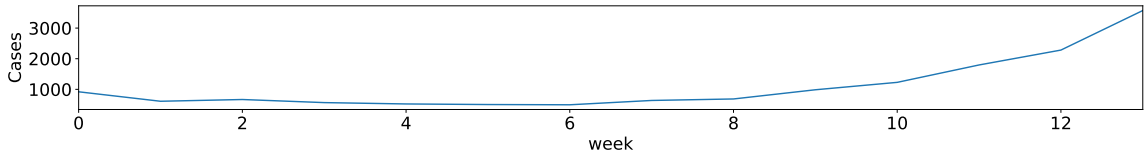


(a) External cases

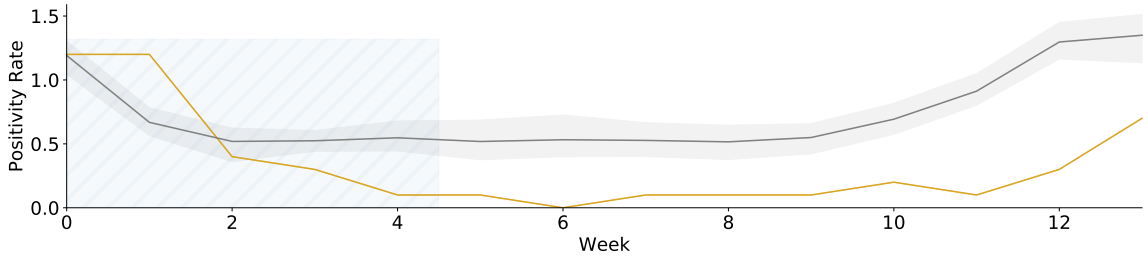


(b) Calibrating on the weeks 0-4

Figure A.2: We calibrate NABM on positivity rates from first 5 weeks of Fall 2020 at UIUC. The objective function of the calibration is to minimize the r.m.s.e. with the weekly average of positivity rate obtained from surveillance testing results at GT [49]. (a) The parameter that determines external transmission of infections on a given day, $I_{out}(t)$, is a function of cases in Champaign county (where UIUC is located). (b) We illustrate the output for a range of parameters that incorporate quantitative uncertainty, i.e., within 40% of the r.m.s.e. The shaded region around the lines show the 2.5^{th} and 97.5^{th} percentile.



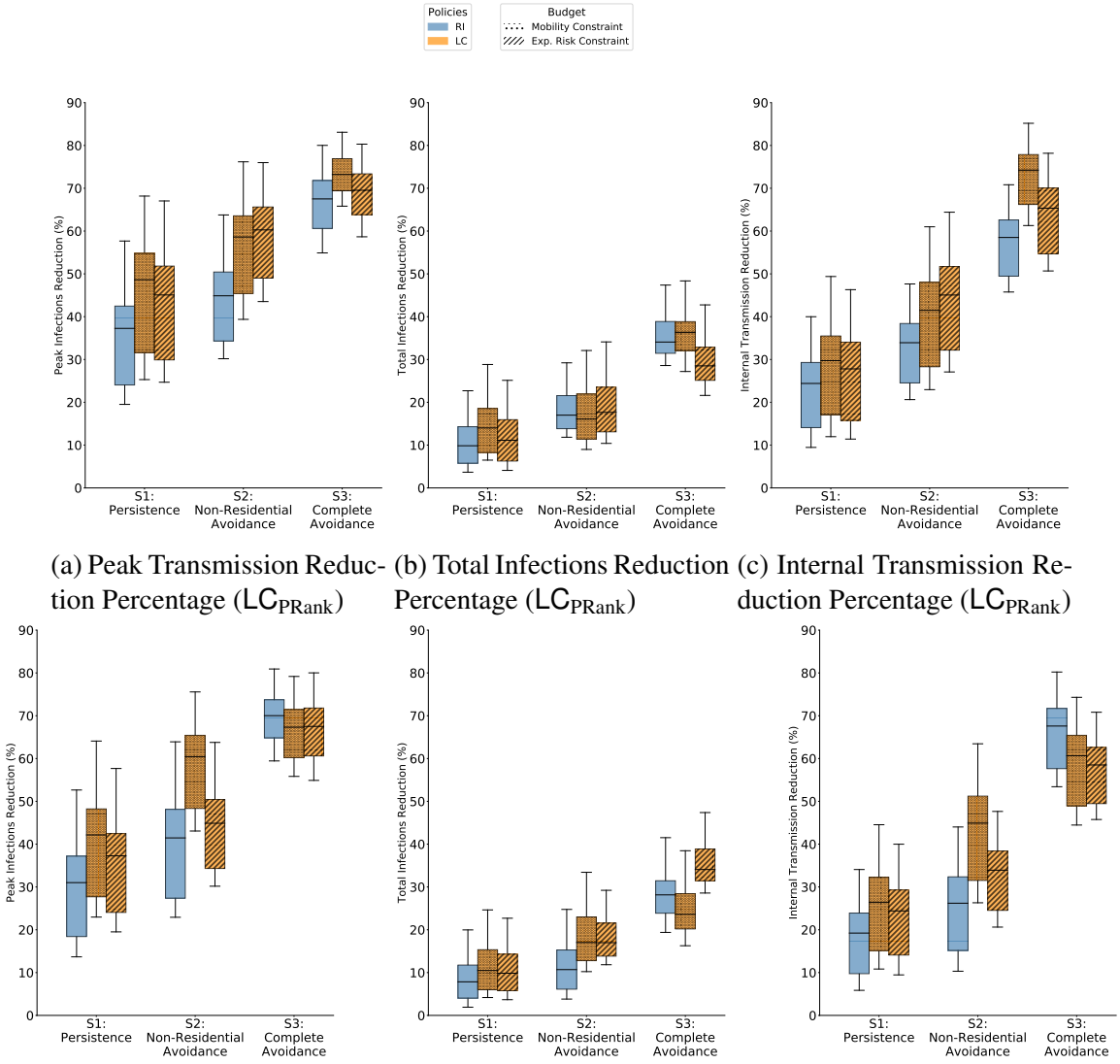
(a) External cases



(b) Calibrating on the weeks 0-4

Figure A.3: Calibration Results (UC Berkeley).

Figure A.4: We calibrate NABM on positivity rates from first 5 weeks of Fall 2020 at UC Berkeley. The objective function of the calibration is to minimize the r.m.s.e. with the weekly average of positivity rate obtained from surveillance testing results at GT [49]. (a) The parameter that determines external transmission of infections on a given day, $I_{out}(t)$, is a function of cases in Alameda county (where UIUC is located). (b) We illustrate the output for a range of parameters that incorporate quantitative uncertainty, i.e., within 40% of the r.m.s.e. The shaded region around the lines show the 2.5^{th} and 97.5^{th} percentile.



(a) Peak Transmission Reduction Percentage ($\mathsf{LC}_{\text{PRank}}$)

(b) Total Infections Reduction Percentage ($\mathsf{LC}_{\text{PRank}}$)

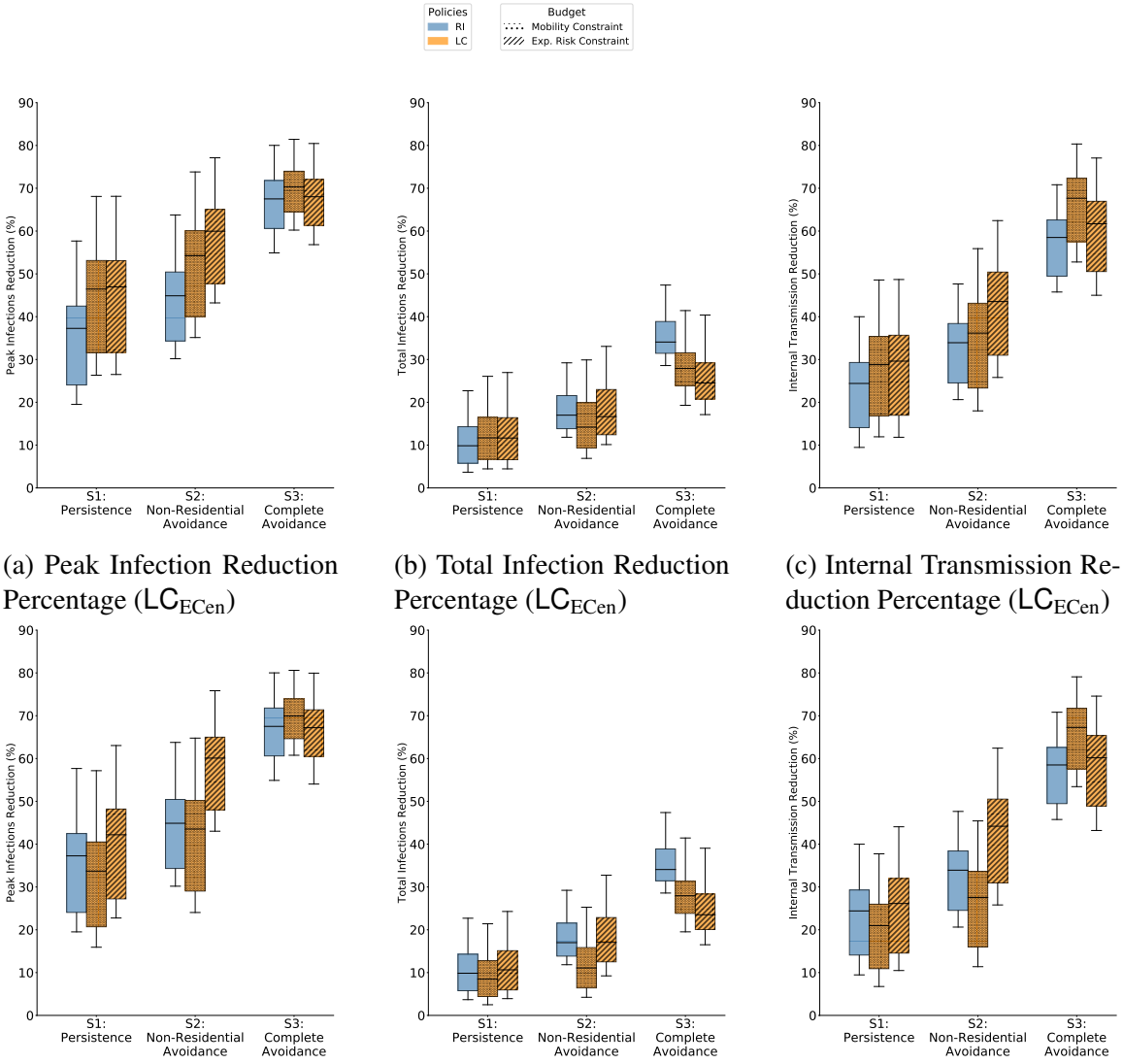
(c) Internal Transmission Reduction Percentage ($\mathsf{LC}_{\text{PRank}}$)

(d) Peak Infection Reduction Percentage ($\mathsf{LC}_{\text{BCen}}$)

(e) Total Infection Reduction Percentage ($\mathsf{LC}_{\text{BCen}}$)

(f) Internal Transmission Reduction Percentage ($\mathsf{LC}_{\text{BCen}}$)

Figure A.5: Disease control outcomes in Fall 2019 for different algorithms of LC with the NABM is calibrated on weeks $0-4$ of Fall 2020 at GT. $(a-c)$ Comparison of RI with $\mathsf{LC}_{\text{PRank}}$. Under all scenarios, for peak infection reduction (b) and internal transmission reduction (c), $\mathsf{LC}_{\text{PRank}}$ shows better disease control outcomes than RI. For total infection reduction (b), $\mathsf{LC}_{\text{PRank}}$ is better in Persistent Scenario (*S1*), worse in *S3* when designed within an exposure risk budget, and comparable in others. $(d-f)$ Comparison of RI with $\mathsf{LC}_{\text{BCen}}$. Under all scenarios, for peak infection reduction (d) and internal transmission reduction (f) $\mathsf{LC}_{\text{BCen}}$ is better when designed within an exposure risk budget. For total infection reduction (e), $\mathsf{LC}_{\text{BCen}}$ is always worse than RI



(a) Peak Infection Reduction Percentage ($\mathsf{LC}_{\text{ECen}}$)

(b) Total Infection Reduction Percentage ($\mathsf{LC}_{\text{ECen}}$)

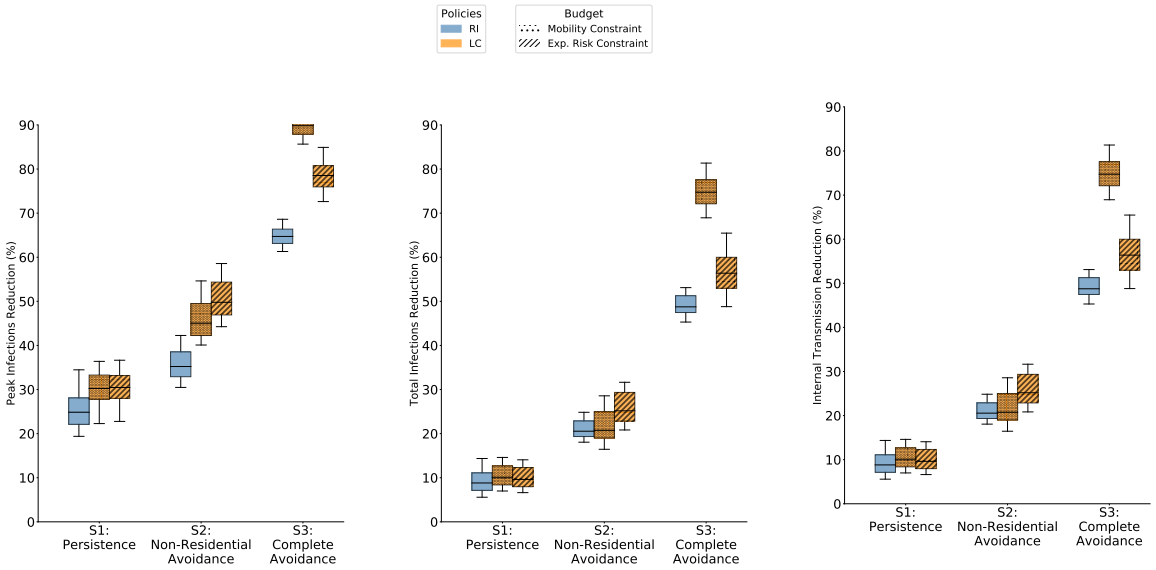
(c) Internal Transmission Reduction Percentage ($\mathsf{LC}_{\text{ECen}}$)

(d) Peak Infection Reduction Percentage ($\mathsf{LC}_{\text{LCen}}$)

(e) Total Infection Reduction Percentage ($\mathsf{LC}_{\text{LCen}}$)

(f) Internal Transmission Reduction Percentage ($\mathsf{LC}_{\text{LCen}}$)

Figure A.6: Disease control outcomes in Fall 2019 for different algorithms of LC with the NABM is calibrated on weeks $0-4$ of Fall 2020 at GT. $(a-c)$ Comparison of RI with $\mathsf{LC}_{\text{ECen}}$. Under all scenarios, for peak infection reduction (b) and internal transmission reduction (c), $\mathsf{LC}_{\text{ECen}}$ shows better disease control outcomes than RI. For total infection reduction (b), $\mathsf{LC}_{\text{ECen}}$ is better in *S1* and worse in *S3*when designed within an exposure risk budget. $(d-f)$ Comparison of RI with $\mathsf{LC}_{\text{ECen}}$. Under all scenarios, for peak infection reduction (d) and internal transmission reduction (f), $\mathsf{LC}_{\text{ECen}}$ shows better disease control outcomes than RI. For total infection reduction (e), $\mathsf{LC}_{\text{ECen}}$ is better in *S1* and worse in *S3*when designed within an exposure risk budget.



(a) Peak Transmission Reduction Percentage (Weeks 5 - 9)

(b) Total Infections Reduction Percentage (Weeks 5 - 9)

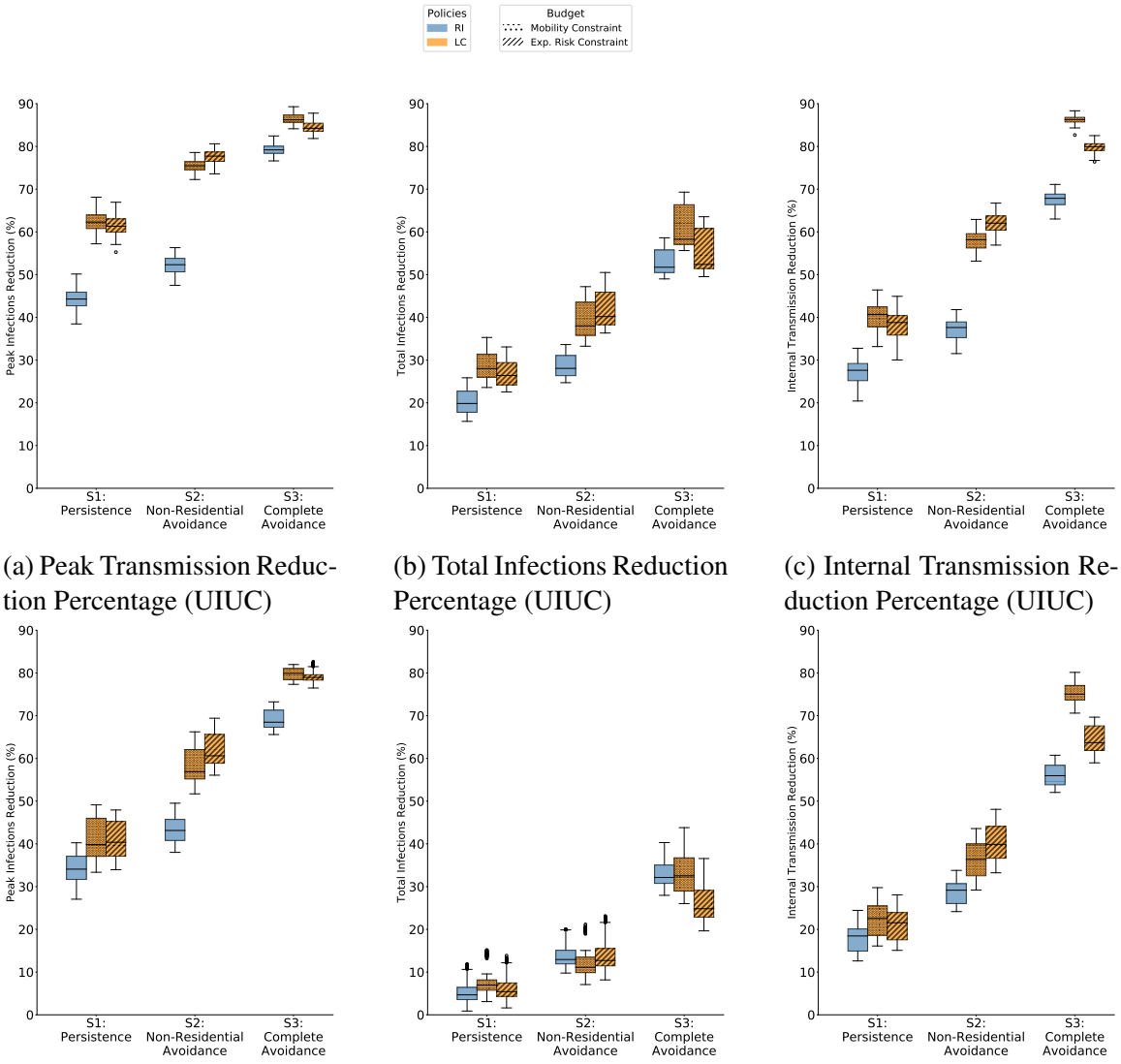
(c) Internal Transmission Reduction Percentage (Weeks 5 - 9)

(d) Peak Transmission Reduction Percentage (Weeks 10 - 14)

(e) Total Infections Reduction Percentage (Weeks 10 - 14)

(f) Internal Transmission Reduction Percentage (Weeks 10 - 14)

Figure A.7: Disease control outcomes in Fall 2019 for $\mathsf{LC}_{\text{PRank}}$. $(a-c)$ The NABM was calibrated on weeks $5-9$ of Fall 2020 at GT. Under all scenarios, for all outcomes, $\mathsf{LC}_{\text{PRank}}$ is better than RI. $(d-f)$ The NABM was calibrated on weeks $10-14$ of Fall 2020 at GT. Under all scenarios, for all outcomes, $\mathsf{LC}_{\text{PRank}}$ is better than RI.



(a) Peak Transmission Reduction Percentage (UIUC)

(b) Total Infections Reduction Percentage (UIUC)

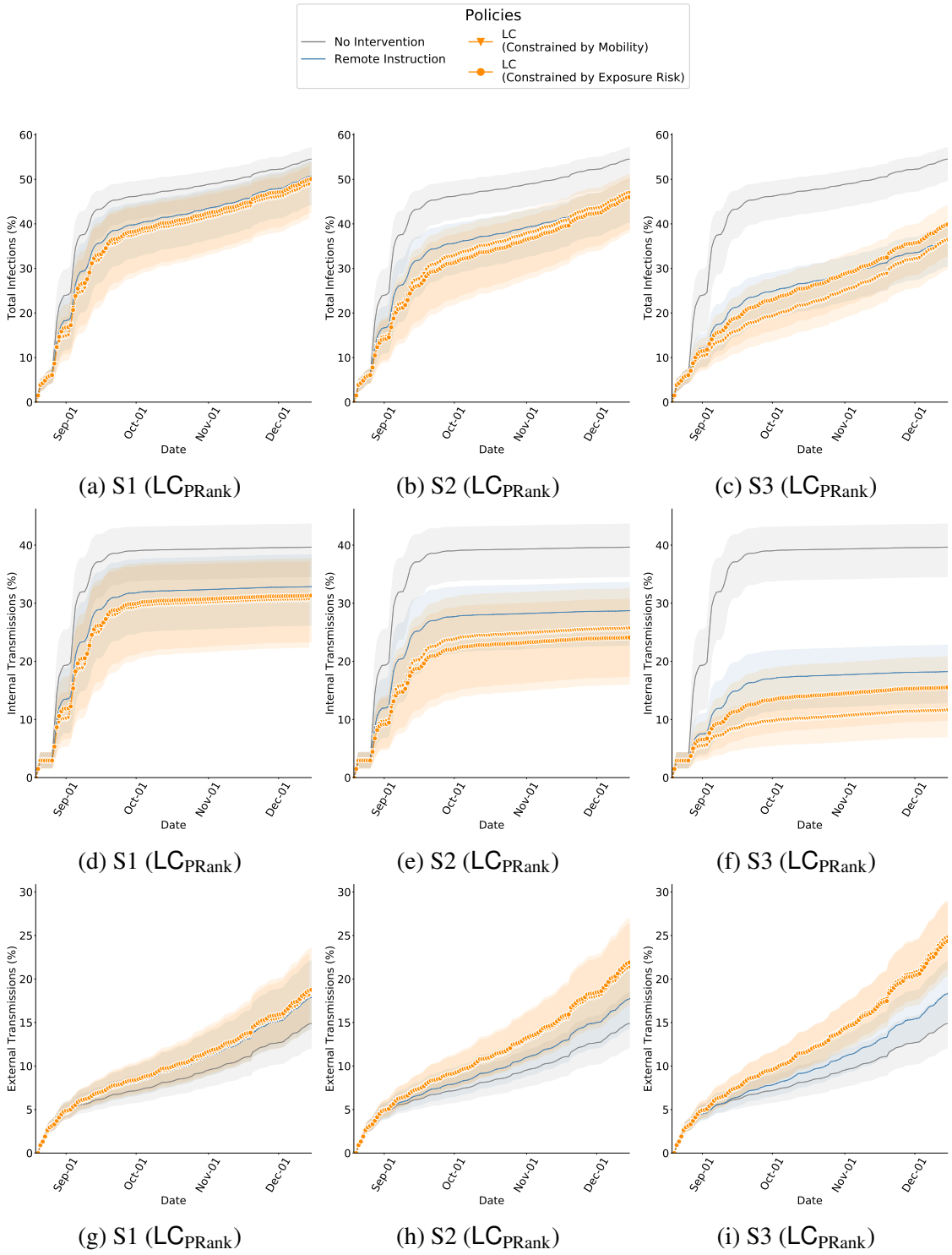
(c) Internal Transmission Reduction Percentage (UIUC)

(d) Peak Transmission Reduction Percentage (UCB)

(e) Total Infections Reduction Percentage (UCB)

(f) Internal Transmission Reduction Percentage (UCB)

Figure A.8: Disease control outcomes in Fall 2019 for $\mathsf{LC}_{\text{PRank}}$. $(a-c)$ The NABM was calibrated on weeks $0-4$ of Fall 2020 at UIUC. Under all scenarios, for all outcomes, $\mathsf{LC}_{\text{PRank}}$ is better than RI. $(d-f)$ The NABM was calibrated on weeks $0-4$ of Fall 2020 at UC Berkeley. Under all scenarios, for all outcomes, $\mathsf{LC}_{\text{PRank}}$ is better than RI.



(a) S1 (LC_{PRank}) (b) S2 (LC_{PRank}) (c) S3 (LC_{PRank})

(d) S1 (LC_{PRank}) (e) S2 (LC_{PRank}) (f) S3 (LC_{PRank})

(g) S1 (LC_{PRank}) (h) S2 (LC_{PRank}) (i) S3 (LC_{PRank})

Figure A.9: Cumulative infections in Fall 2019 while comparing RI and LC_{PRank} with NABM calibrated on weeks $0-4$ of Fall 2020, GT. The bands show the 2.75^{th} and 97.25^{th} percentile. $(a-c)$ Total infections of interventions is lower than no-intervention scenarios and is lowest in the *S3* scenario. In this scenario, the mobility budget is 69% of what it would be without interventions, and therefore the transmissions are also contained. In comparison, in Fall 2020, we saw far fewer infections which is because the mobility was 39% of that in Fall 2019. $(d-f)$ Internal transmissions are lower with LC_{PRank} in comparison to RI. $(g-i)$ External transmissions are higher with LC_{PRank} in comparison to RI. Since internal transmission is controlled, more individuals remain susceptible to infections from outside campus.

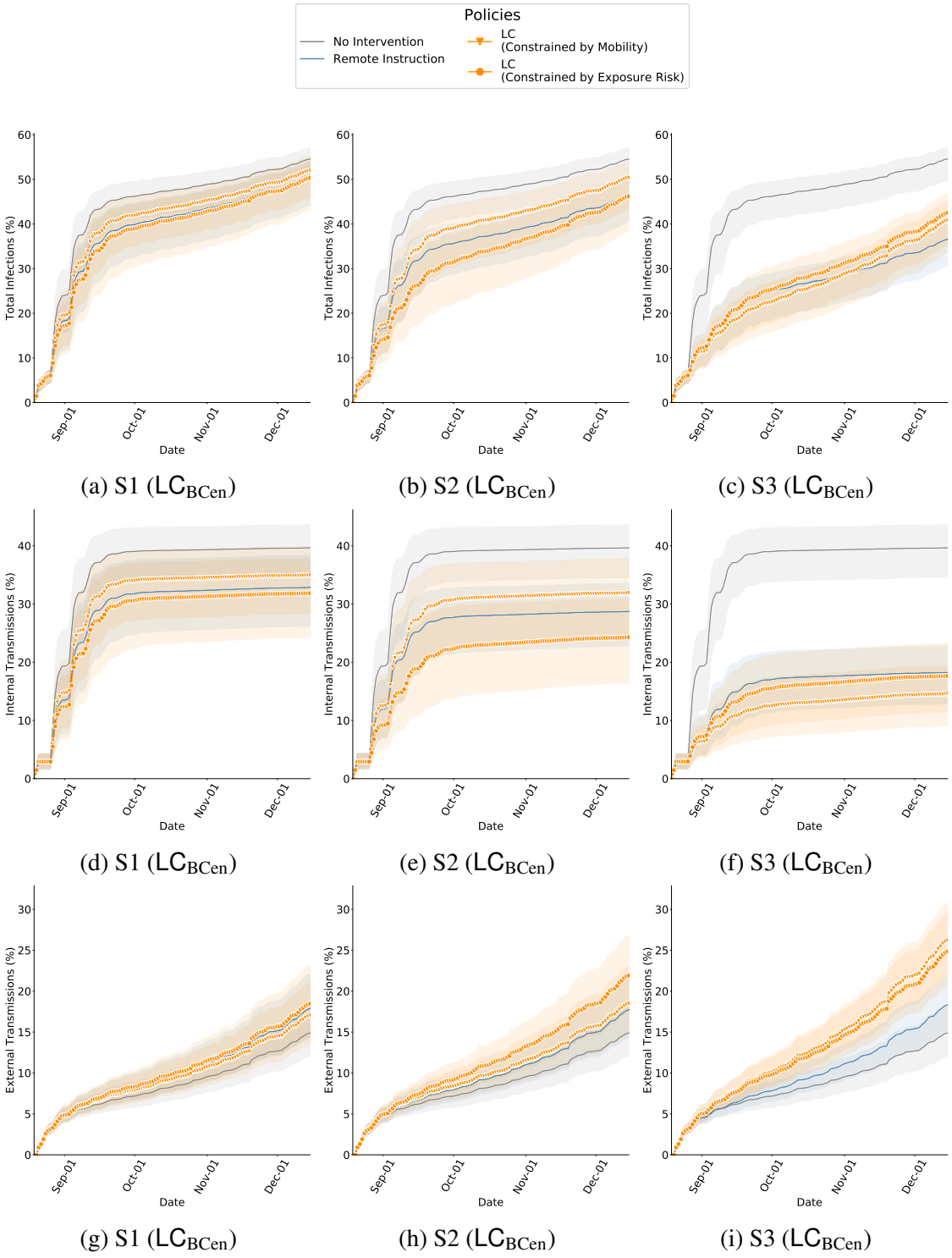
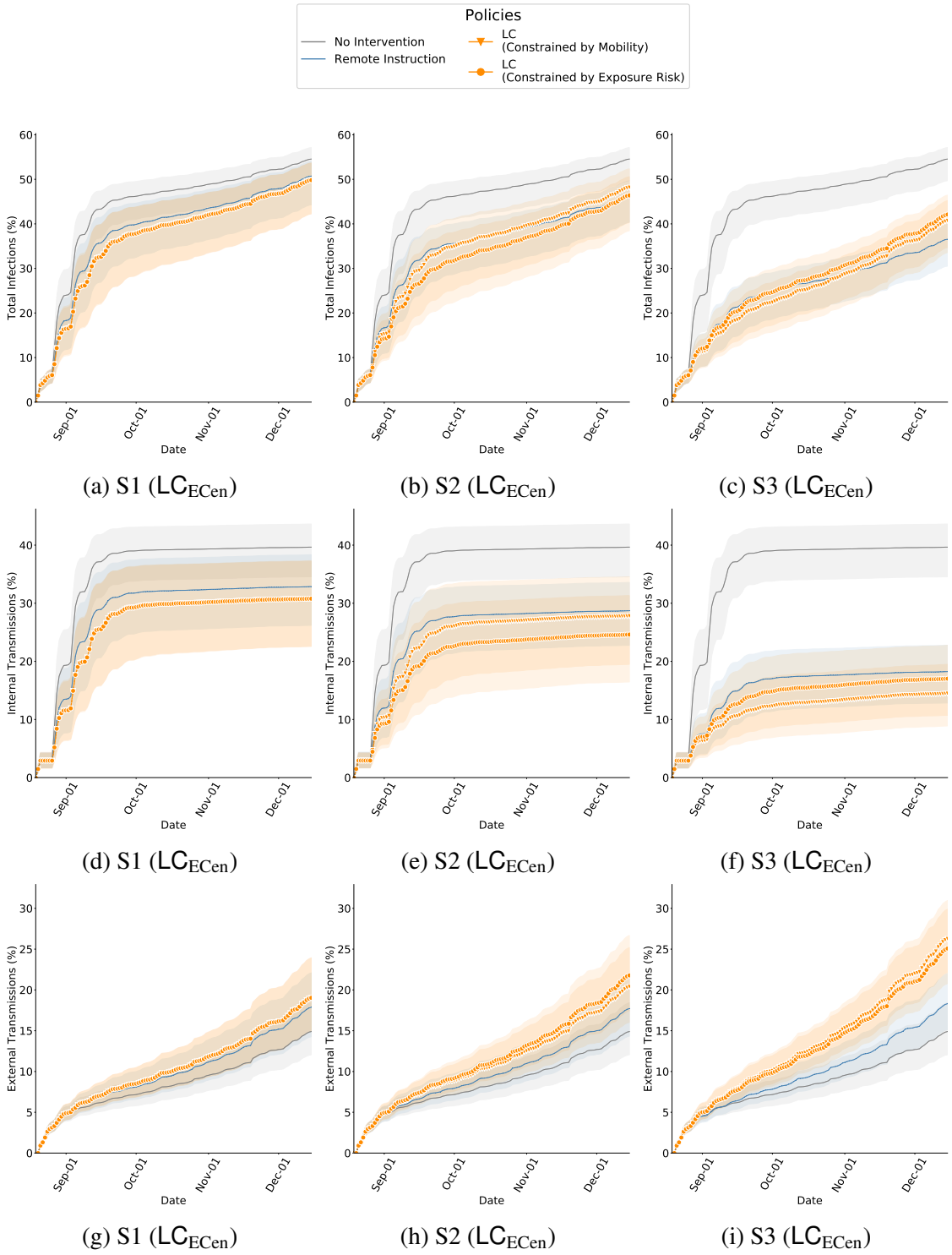


Figure A.10: Cumulative infections in Fall 2019 while comparing RI and LC$_{BCen}$ with NABM calibrated on weeks $0-4$ of Fall 2020, GT. The bands show the 2.75^{th} and 97.25^{th} percentile. $(a-c)$ Total infections of interventions is lower than no-intervention scenarios and is lowest in the *S3* scenario. In this scenario, the mobility budget is 69% of what it would be without interventions, and therefore the transmissions are also contained. In comparison, in Fall 2020, we saw far fewer infections which is because the mobility was 39% of that in Fall 2019. $(d-f)$ Internal transmissions are lower with LC$_{BCen}$ in comparison to RI, only when constrained under the exposure risk budget. $(g-i)$ External transmissions are higher with LC$_{BCen}$ in comparison to RI. Since internal transmission is controlled, more individuals remain susceptible to infections from outside campus.

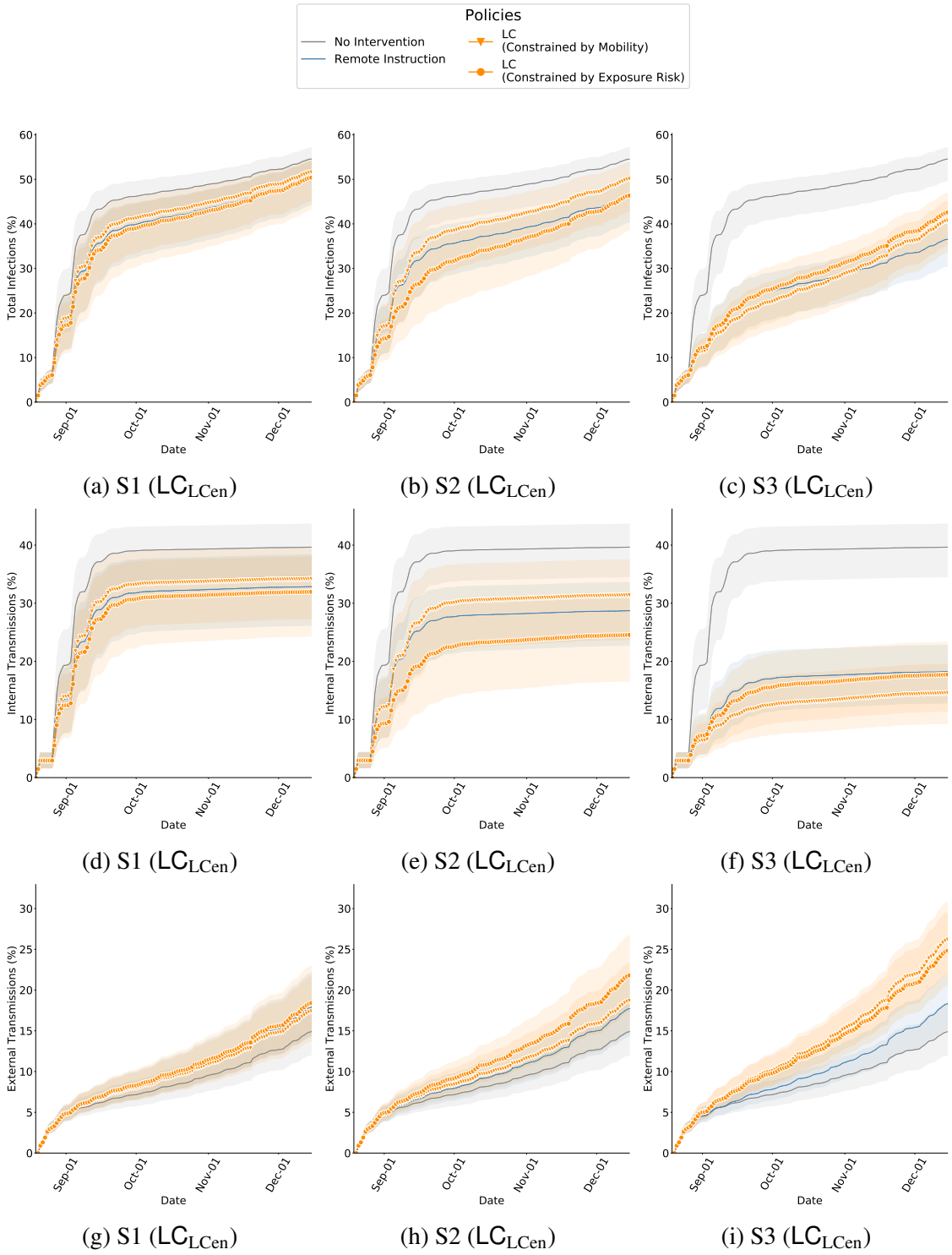


(a) S1 (LC_{ECen}) (b) S2 (LC_{ECen}) (c) S3 (LC_{ECen})

(d) S1 (LC_{ECen}) (e) S2 (LC_{ECen}) (f) S3 (LC_{ECen})

(g) S1 (LC_{ECen}) (h) S2 (LC_{ECen}) (i) S3 (LC_{ECen})

Figure A.11: Cumulative infections in Fall 2019 while comparing RI and LC_{ECen} with NABM calibrated on weeks $0-4$ of Fall 2020, GT. The bands show the 2.75^{th} and 97.25^{th} percentile. $(a-c)$ Total infections of interventions is lower than no-intervention scenarios and is lowest in the *S3* scenario. In this scenario, the mobility budget is 69% of what it would be without interventions, and therefore the transmissions are also contained. In comparison, in Fall 2020, we saw far fewer infections which is because the mobility was 39% of that in Fall 2019. $(d-f)$ Internal transmissions are lower with LC_{ECen} in comparison to RI. $(g-i)$ External transmissions are higher with LC_{ECen} in comparison to RI. Since internal transmission is controlled, more individuals remain susceptible to infections from outside campus.

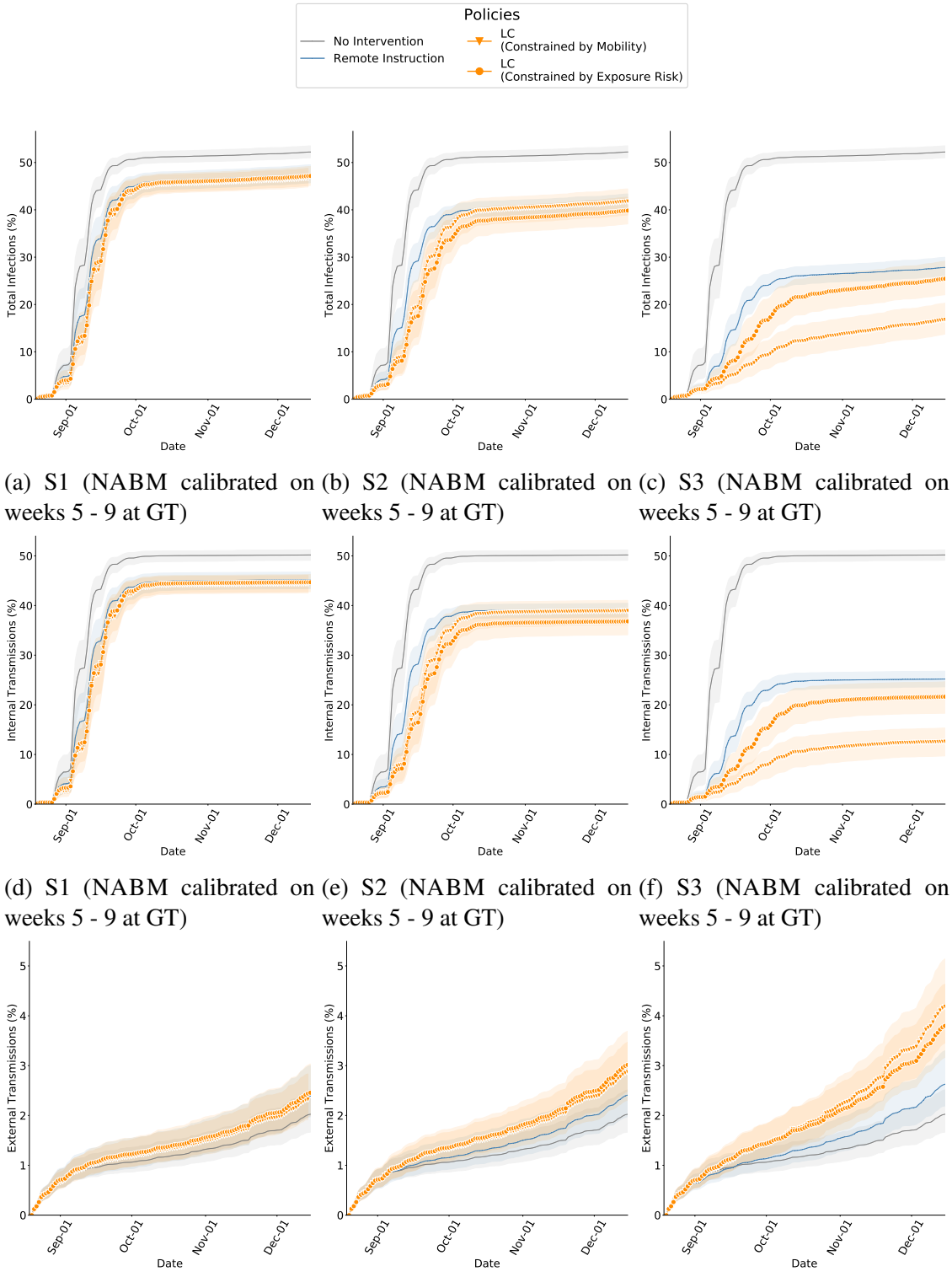


(a) S1 ($\mathsf{LC}_{\mathrm{LCen}}$) (b) S2 ($\mathsf{LC}_{\mathrm{LCen}}$) (c) S3 ($\mathsf{LC}_{\mathrm{LCen}}$)

(d) S1 ($\mathsf{LC}_{\mathrm{LCen}}$) (e) S2 ($\mathsf{LC}_{\mathrm{LCen}}$) (f) S3 ($\mathsf{LC}_{\mathrm{LCen}}$)

(g) S1 ($\mathsf{LC}_{\mathrm{LCen}}$) (h) S2 ($\mathsf{LC}_{\mathrm{LCen}}$) (i) S3 ($\mathsf{LC}_{\mathrm{LCen}}$)

Figure A.12: Cumulative infections in Fall 2019 while comparing RI and $\mathsf{LC}_{\mathrm{LCen}}$ with NABM calibrated on weeks $0-4$ of Fall 2020, GT. The bands show the 2.75^{th} and 97.25^{th} percentile. $(a-c)$ Total infections of interventions is lower than no-intervention scenarios and is lowest in the *S3* scenario. In this scenario, the mobility budget is 69% of what it would be without interventions, and therefore the transmissions are also contained. In comparison, in Fall 2020, we saw far fewer infections which is because the mobility was 39% of that in Fall 2019. $(d-f)$ Internal transmissions are lower with $\mathsf{LC}_{\mathrm{LCen}}$ in comparison to RI. $(g-i)$ External transmissions are higher with $\mathsf{LC}_{\mathrm{LCen}}$ in comparison to RI. Since internal transmission is controlled, more individuals remain susceptible to infections from outside campus.



(a) S1 (NABM calibrated on weeks 5 - 9 at GT)

(b) S2 (NABM calibrated on weeks 5 - 9 at GT)

(c) S3 (NABM calibrated on weeks 5 - 9 at GT)

(d) S1 (NABM calibrated on weeks 5 - 9 at GT)

(e) S2 (NABM calibrated on weeks 5 - 9 at GT)

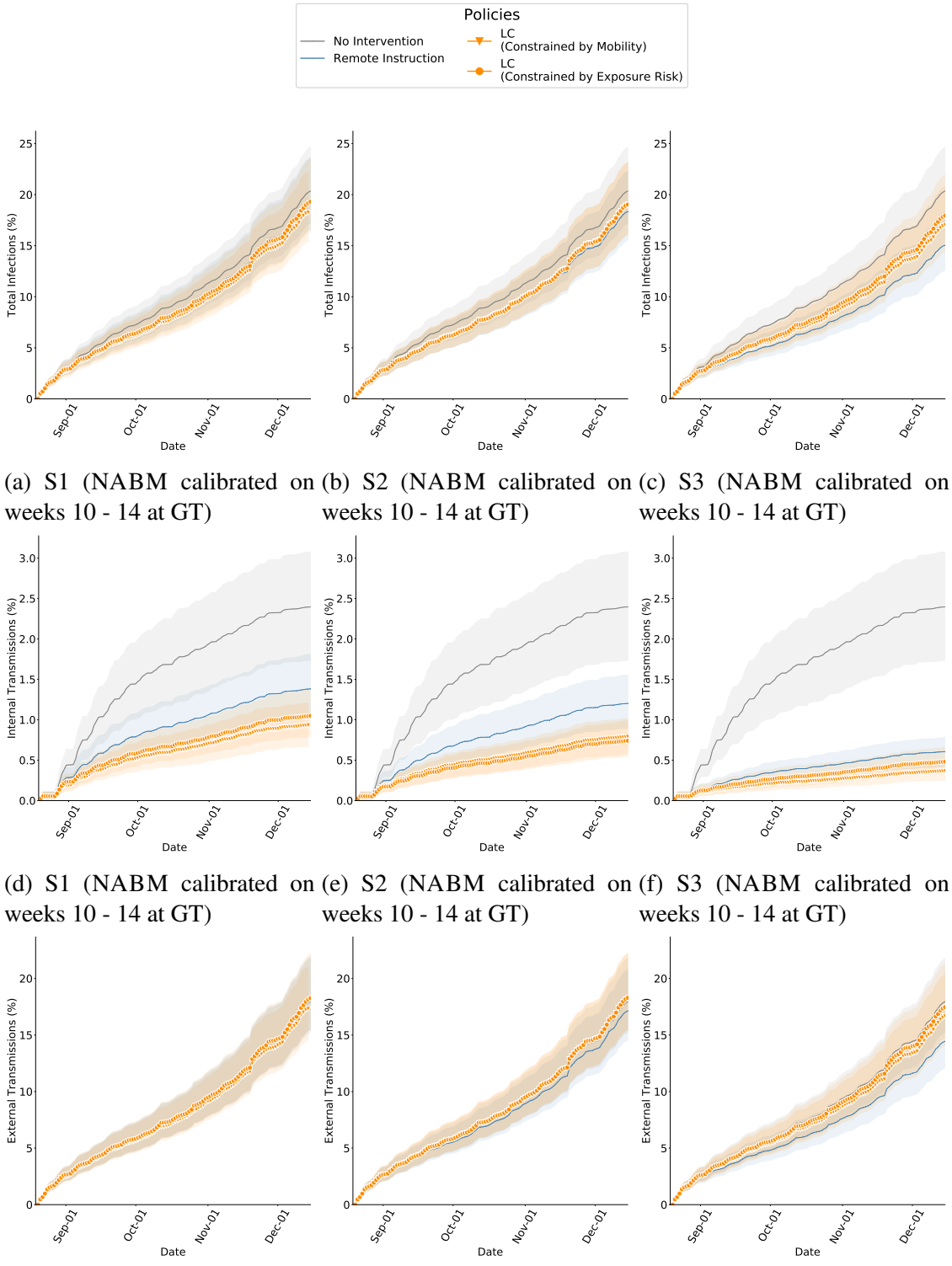
(f) S3 (NABM calibrated on weeks 5 - 9 at GT)

(g) S1 (NABM calibrated on weeks 5 - 9 at GT)

(h) S2 (NABM calibrated on weeks 5 - 9 at GT)

(i) S3 (NABM calibrated on weeks 5 - 9 at GT)

Figure A.13: Cumulative infections in Fall 2019 while comparing RI and LC_{PRank} with NABM calibrated on weeks $5-9$ of Fall 2020, GT. The bands show the 2.75^{th} and 97.25^{th} percentile. $(a-c)$ Total infections of interventions is lower than no-intervention scenarios and is lowest in the *S3* scenario. In this scenario, the mobility budget is 69% of what it would be without interventions, and therefore the transmissions are also contained. In comparison, in Fall 2020, we saw far fewer infections which is because the mobility was 39% of that in Fall 2019. $(d-f)$ Internal transmissions are lower with LC_{PRank} in comparison to RI. $(g-i)$ External transmissions are higher with LC_{PRank} in comparison to RI. Since internal transmission is controlled, more individuals remain susceptible to infections from outside campus.



(a) S1 (NABM calibrated on weeks 10 - 14 at GT)

(b) S2 (NABM calibrated on weeks 10 - 14 at GT)

(c) S3 (NABM calibrated on weeks 10 - 14 at GT)

(d) S1 (NABM calibrated on weeks 10 - 14 at GT)

(e) S2 (NABM calibrated on weeks 10 - 14 at GT)

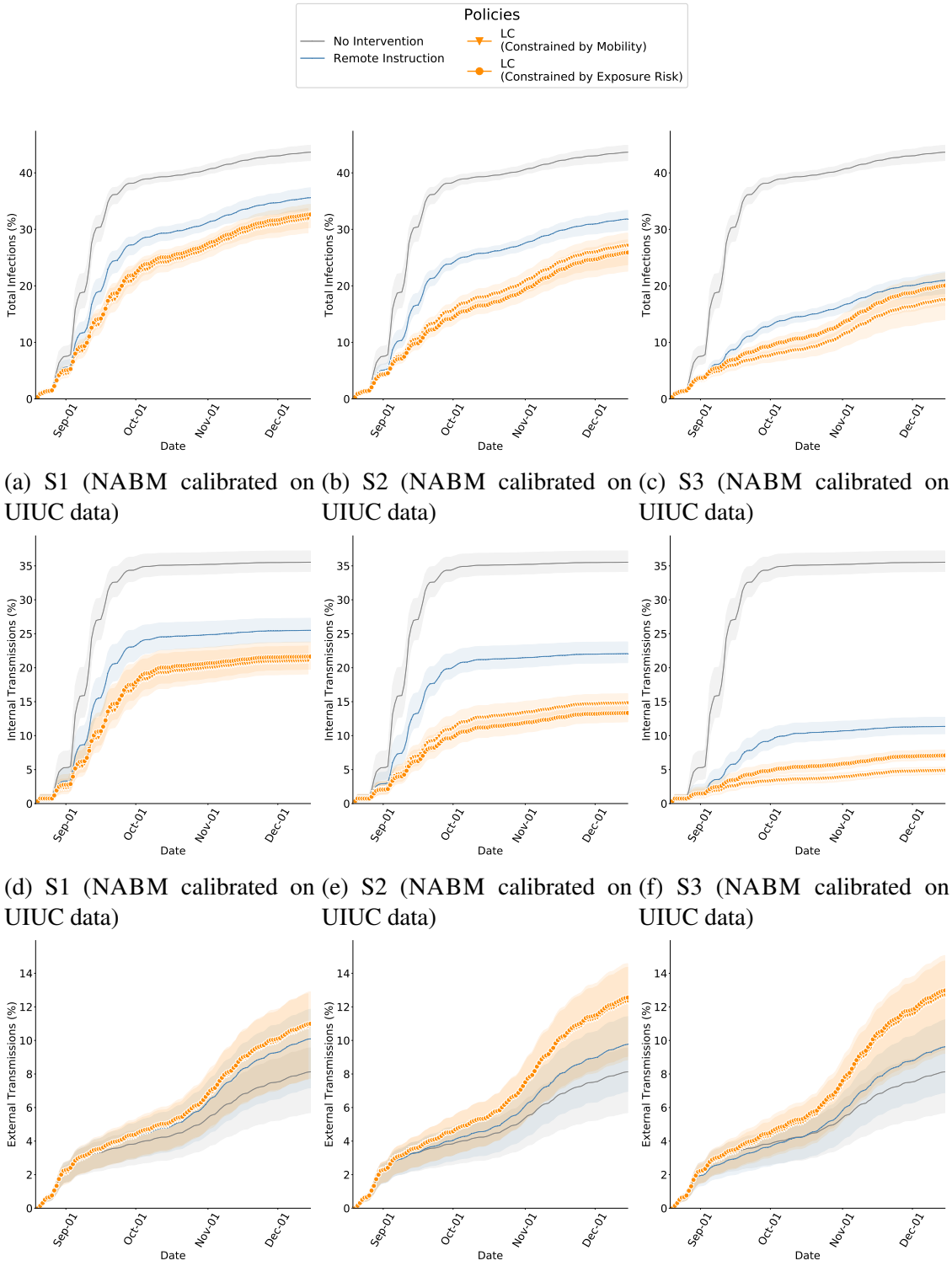
(f) S3 (NABM calibrated on weeks 10 - 14 at GT)

(g) S1 (NABM calibrated on weeks 10 - 14 at GT)

(h) S2 (NABM calibrated on weeks 10 - 14 at GT)

(i) S3 (NABM calibrated on weeks 10 - 14 at GT)

Figure A.14: Cumulative infections in Fall 2019 while comparing RI and LC_{PRank} with NABM calibrated on weeks $10 - 14$ of Fall 2020, GT. The bands show the 2.75^{th} and 97.25^{th} percentile. $(a - c)$ Total infections of interventions is lower than no-intervention scenarios and is lowest in the *S3* scenario. In this scenario, the mobility budget is 69% of what it would be without interventions, and therefore the transmissions are also contained. In comparison, in Fall 2020, we saw far fewer infections which is because the mobility was 39% of that in Fall 2019. $(d - f)$ Internal transmissions are lower with LC_{PRank} in comparison to RI. $(g - i)$ External transmissions are higher with LC_{PRank} in comparison to RI. Since internal transmission is controlled, more individuals remain susceptible to infections from outside campus.



(a) S1 (NABM calibrated on UIUC data)

(b) S2 (NABM calibrated on UIUC data)

(c) S3 (NABM calibrated on UIUC data)

(d) S1 (NABM calibrated on UIUC data)

(e) S2 (NABM calibrated on UIUC data)

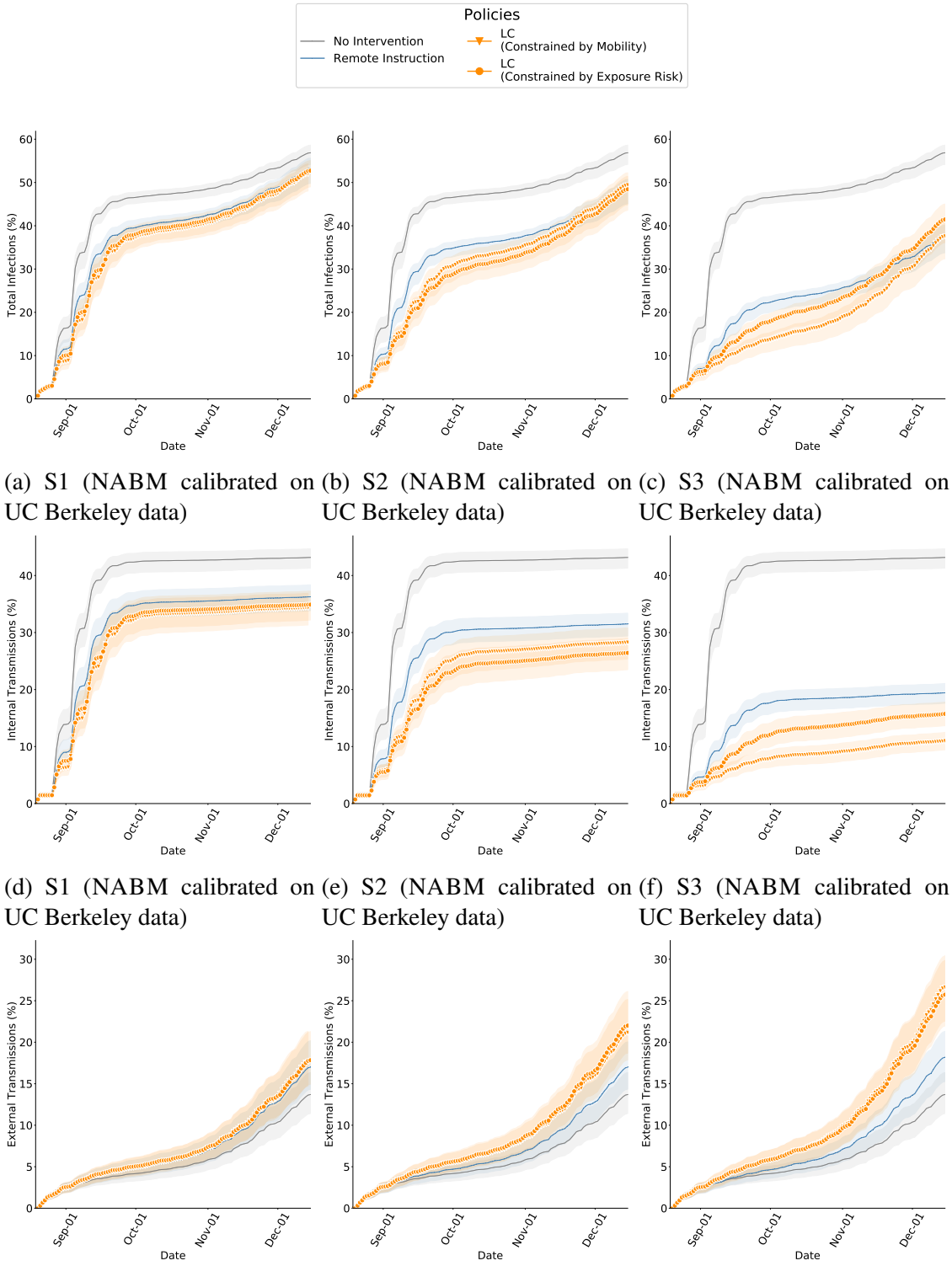
(f) S3 (NABM calibrated on UIUC data)

(g) S1 (NABM calibrated on UIUC data)

(h) S2 (NABM calibrated on UIUC data)

(i) S3 (NABM calibrated on UIUC data)

Figure A.15: Cumulative infections in Fall 2019 while comparing RI and LC_{PRank} with NABM calibrated on weeks $0-4$ of Fall 2020, UIUC. The bands show the 2.75^{th} and 97.25^{th} percentile. $(a-c)$ Total infections of interventions is lower than no-intervention scenarios and is lowest in the *S3* scenario. In this scenario, the mobility budget is 69% of what it would be without interventions, and therefore the transmissions are also contained. In comparison, in Fall 2020, we saw far fewer infections which is because the mobility was 39% of that in Fall 2019. $(d-f)$ Internal transmissions are lower with LC_{PRank} in comparison to RI. $(g-i)$ External transmissions are higher with LC_{PRank} in comparison to RI. Since internal transmission is controlled, more individuals remain susceptible to infections from outside campus.



(a) S1 (NABM calibrated on UC Berkeley data)

(b) S2 (NABM calibrated on UC Berkeley data)

(c) S3 (NABM calibrated on UC Berkeley data)

(d) S1 (NABM calibrated on UC Berkeley data)

(e) S2 (NABM calibrated on UC Berkeley data)

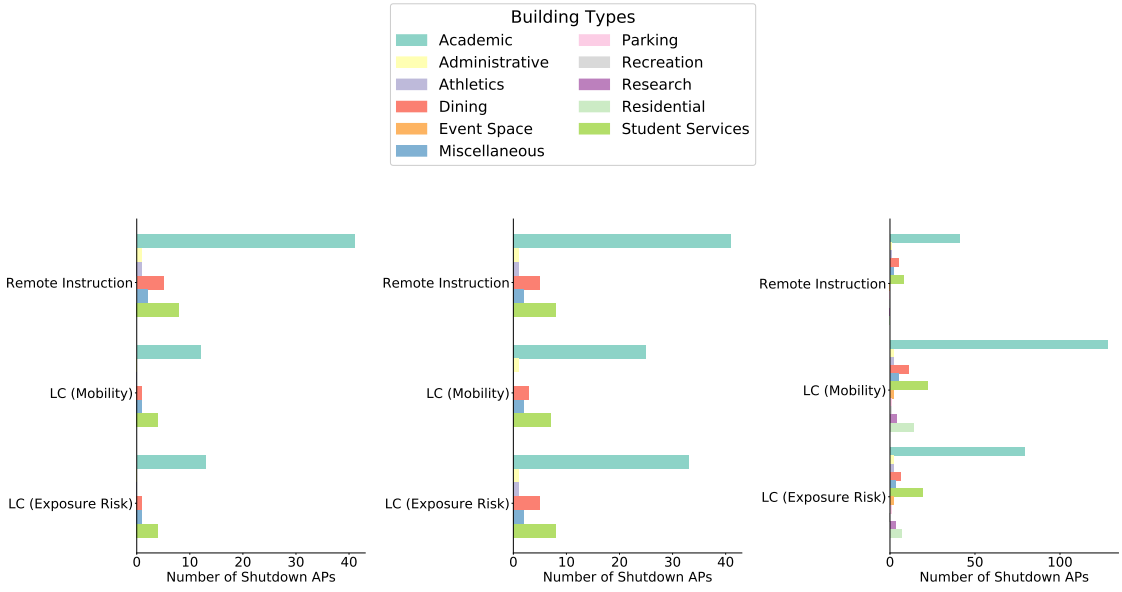
(f) S3 (NABM calibrated on UC Berkeley data)

(g) S1 (NABM calibrated on UC Berkeley data)

(h) S2 (NABM calibrated on UC Berkeley data)

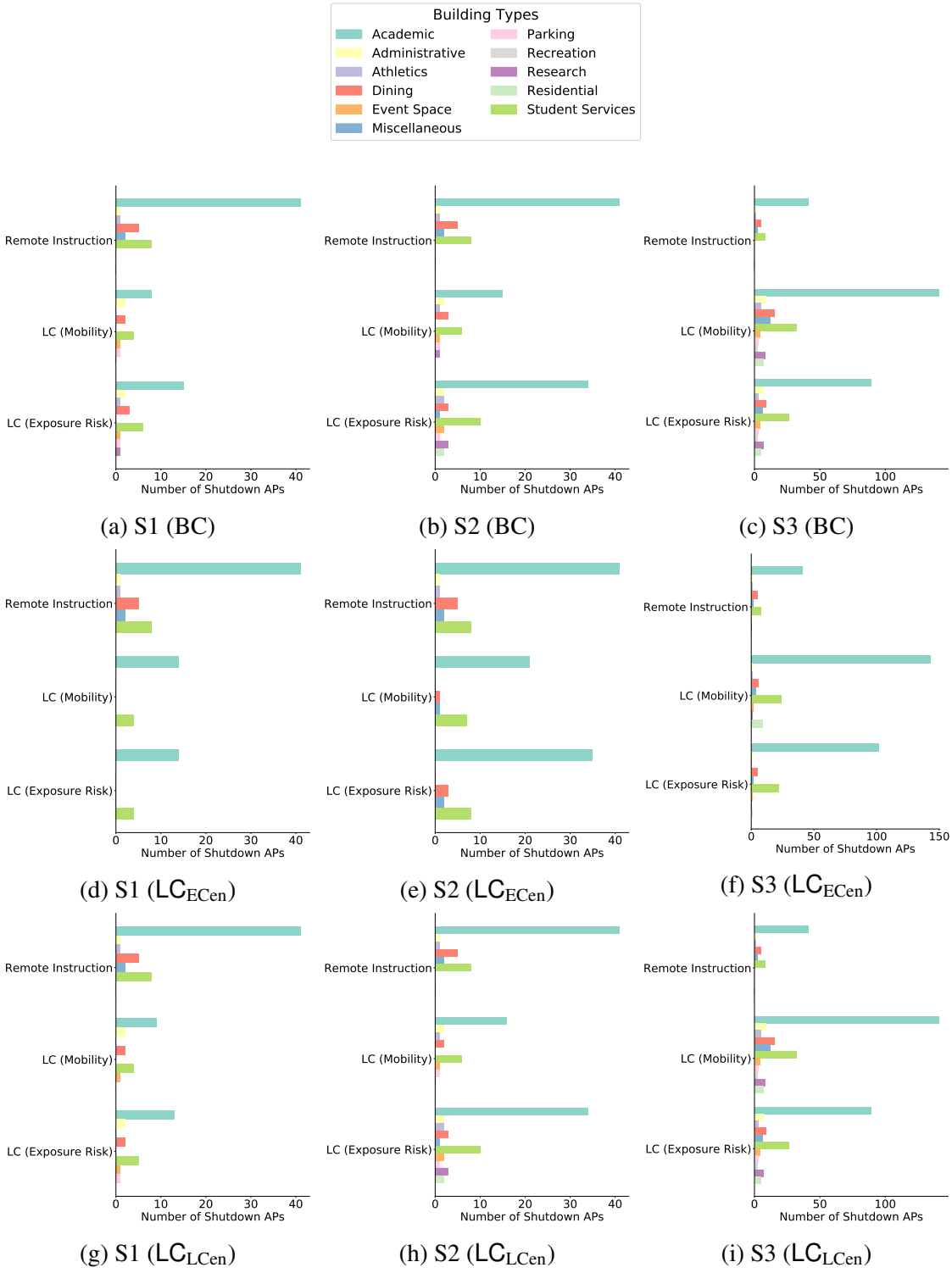
(i) S3 (NABM calibrated on UC Berkeley data)

Figure A.16: Cumulative infections in Fall 2019 while comparing RI and LC_{PRank} with NABM calibrated on weeks $0-4$ of Fall 2020, UC Berkeley. The bands show the 2.75^{th} and 97.25^{th} percentile. $(a-c)$ Total infections of interventions is lower than no-intervention scenarios and is lowest in the *S3* scenario. In this scenario, the mobility budget is 69% of what it would be without interventions, and therefore the transmissions are also contained. In comparison, in Fall 2020, we saw far fewer infections which is because the mobility was 39% of that in Fall 2019. $(d-f)$ Internal transmissions are lower with LC_{PRank} in comparison to RI. $(g-i)$ External transmissions are higher with LC_{PRank} in comparison to RI. Since internal transmission is controlled, more individuals remain susceptible to infections from outside campus.



(a) S1 ($\mathsf{LC}_{\text{PRank}}$) (b) S2 ($\mathsf{LC}_{\text{PRank}}$) (c) S3 ($\mathsf{LC}_{\text{PRank}}$)

Figure A.17: The locations shutdown by each policy are grouped into the the general building category. The distribution of locations is different between policies, for example, in *S1* (a) and Non-residential Avoidance (*S2*) (b), LC closes fewer locations that RI. Even when targeting spaces in similar buildings, the locations are qualitatively different — RI only affects classrooms, whereas LC also closes smaller spaces like breakout rooms, reading areas and cafes. LC In *S3* (c) we find LC to target locations in a greater variety of buildings, but it also targets more locations to utilize the budget.



(a) S1 (BC) (b) S2 (BC) (c) S3 (BC)

(d) S1 (LC_{ECen}) (e) S2 (LC_{ECen}) (f) S3 (LC_{ECen})

(g) S1 (LC_{LCen}) (h) S2 (LC_{LCen}) (i) S3 (LC_{LCen})

Figure A.18: The locations shutdown by each policy are grouped into the the general building category. The distribution of locations is different between policies, for example, in *S1* (a) and *S2* (b), LC closes fewer locations that RI. Even when targeting spaces in similar buildings, the locations are qualitatively different — RI only affects classrooms, whereas LC also closes smaller spaces like breakout rooms, reading areas and cafes. LC In *S3* (c) we find LC to target locations in a greater variety of buildings, but it also targets more locations to utilize the budget.

REFERENCES

[1] C. for Disease Control, P. (CDC, *et al.*, "Mumps outbreak on a university campus–california, 2011," *MMWR. Morbidity and mortality weekly report*, vol. 61, no. 48, pp. 986–989, 2012.

[2] R. W. Amler, R. J. Kim-Farley, W. A. Orenstein, S. W. Doster, and K. J. Bart, "Measles on campus," *Journal of American College Health*, vol. 32, no. 2, pp. 53–57, 1983.

[3] W. H. Organization *et al.*, "Critical preparedness, readiness and response actions for covid-19: Interim guidance, 22 march 2020," World Health Organization, Tech. Rep., 2020.

[4] E. Wilson, C. V. Donovan, M. Campbell, T. Chai, K. Pittman, A. C. Seña, A. Pettifor, D. J. Weber, A. Mallick, A. Cope, *et al.*, "Multiple covid-19 clusters on a university campus—north carolina, august 2020," *Morbidity and Mortality Weekly Report*, vol. 69, no. 39, p. 1416, 2020.

[5] E. Mahmud, H. L. Dauerman, F. G. Welt, J. C. Messenger, S. V. Rao, C. Grines, A. Mattu, A. J. Kirtane, R. Jauhar, P. Meraj, *et al.*, "Management of acute myocardial infarction during the covid-19 pandemic," *Journal of the American College of Cardiology*, 2020.

[6] M. Harris and K. Holley, "Universities as anchor institutions: Economic and social potential for urban development," in *Higher education: Handbook of theory and research*, Springer, 2016, pp. 393–439.

[7] S. Watson, S. Hubler, D. Ivory, and R. Gebeloff, *A new front in america's pandemic: College towns*, Available at: https://www.nytimes.com/2020/09/06/us/colleges-coronavirus-students.html, 2020.

[8] N. Hasan and Y. Bao, "Impact of "e-learning crack-up" perception on psychological distress among college students during covid-19 pandemic: A mediating role of "fear of academic year loss"," *Children and Youth Services Review*, vol. 118, p. 105 355, 2020.

[9] L. Murphy, N. B. Eduljee, and K. Croteau, "College student transition to synchronous virtual classes during the covid-19 pandemic in northeastern united states.," *Pedagogical Research*, vol. 5, no. 4, 2020.

[10] B. Lopman, C. Y. Liu, A. Le Guillou, A. Handel, T. L. Lash, A. P. Isakov, and S. M. Jenness, "A modeling study to inform screening and testing interventions for

the control of sars-cov-2 on university campuses," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[11] A. D. Paltiel, A. Zheng, and R. P. Walensky, "Assessment of sars-cov-2 screening strategies to permit the safe reopening of college campuses in the united states," *JAMA network open*, vol. 3, no. 7, e2016818–e2016818, 2020.

[12] J. M. Cashore, N. Duan, A. Janmohamed, J. Wan, Y. Zhang, S. Henderson, D. Shmoys, and P. Frazier, "Covid-19 mathematical modeling for cornell's fall semester (2020)," *URL https://people. orie. cornell. edu/pfrazier/COVID*, vol. 19,

[13] I. Holmdahl and C. Buckee, "Wrong but useful—what covid-19 epidemiologic models can and cannot tell us," *New England Journal of Medicine*, vol. 383, no. 4, pp. 303–305, 2020.

[14] R. Bahl, N. Eikmeier, A. Fraser, M. Junge, F. Keesing, K. Nakahata, and L. Z. Wang, "Modeling covid-19 spread in small colleges," *arXiv preprint arXiv:2008.09597*, 2020.

[15] E. M. Hill, B. D. Atkins, M. J. Keeling, M. Tildesley, and L. Dyson, "Modelling sars-cov-2 transmission in a uk university setting," *medRxiv*, 2020.

[16] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the national academy of sciences*, vol. 99, no. suppl 3, pp. 7280–7287, 2002.

[17] K. A. Weeden and B. Cornwell, "The small-world network of college classes: Implications for epidemic spread on a university campus," *Sociological science*, vol. 7, pp. 222–241, 2020.

[18] P. T. Gressman and J. R. Peck, "Simulating covid-19 in a university environment," *Mathematical biosciences*, vol. 328, p. 108 436, 2020.

[19] M. Borowiak, F. Ning, J. Pei, S. Zhao, H.-R. Tung, and R. Durrett, "Controlling the spread of covid-19 on college campuses," *arXiv preprint arXiv:2008.07293*, 2020.

[20] M. Garner, S. Hamilton, *et al.*, "Principles of epidemiological modelling," *Revue Scientifique et Technique-OIE*, vol. 30, no. 2, p. 407, 2011.

[21] V. Subramanian and M. W. Kattan, "Why is modeling coronavirus disease 2019 so difficult?" *Chest*, vol. 158, no. 5, p. 1829, 2020.

[22] P. Holme, "Model versions and fast algorithms for network epidemiology," *arXiv preprint arXiv:1403.1011*, 2014.

[23] L. Meyers, "Contact network epidemiology: Bond percolation applied to infectious disease prediction and control," *Bulletin of the American Mathematical Society*, vol. 44, no. 1, pp. 63–86, 2007.

[24] K. Dietz and J. Heesterbeek, "Daniel bernoulli's epidemiological model revisited," *Mathematical biosciences*, vol. 180, no. 1-2, pp. 1–21, 2002.

[25] M. Mandal, S. Jana, S. K. Nandi, A. Khatua, S. Adak, and T. Kar, "A model based study on the dynamics of covid-19: Prediction and control," *Chaos, Solitons & Fractals*, vol. 136, p. 109 889, 2020.

[26] S. Wan, Y. Xiang, W. Fang, Y. Zheng, B. Li, Y. Hu, C. Lang, D. Huang, Q. Sun, Y. Xiong, *et al.*, "Clinical features and treatment of covid-19 patients in northeast chongqing," *Journal of medical virology*, vol. 92, no. 7, pp. 797–806, 2020.

[27] W. Song, J. Li, N. Zou, W. Guan, J. Pan, and W. Xu, "Clinical features of pediatric patients with coronavirus disease (covid-19)," *Journal of Clinical Virology*, vol. 127, p. 104 377, 2020.

[28] F. Brauer, "Compartmental models in epidemiology," in *Mathematical epidemiology*, Springer, 2008, pp. 19–79.

[29] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM review*, vol. 42, no. 4, pp. 599–653, 2000.

[30] C. McAloon, Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, *et al.*, "Incubation period of covid-19: A rapid systematic review and meta-analysis of observational research," *BMJ open*, vol. 10, no. 8, e039652, 2020.

[31] B. Wilder, M. Charpignon, J. A. Killian, H.-C. Ou, A. Mate, S. Jabbari, A. Perrault, A. N. Desai, M. Tambe, and M. S. Majumder, "Modeling between-population variation in covid-19 dynamics in hubei, lombardy, and new york city," *Proceedings of the National Academy of Sciences*, vol. 117, no. 41, pp. 25 904–25 910, 2020.

[32] P. Keskinocak, B. E. Oruc, A. Baxter, J. Asplund, and N. Serban, "The impact of social distancing on covid19 spread: State of georgia case study," *Plos one*, vol. 15, no. 10, e0239798, 2020.

[33] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec, "Mobility network models of covid-19 explain inequities and inform reopening," *Nature*, vol. 589, no. 7840, pp. 82–87, 2021.

[34] Cisco, *Wi-fi location-based services 4.1 design guide white paper*, Accessed: 2020-05-10, 2014.

[35] A. W.-F. L. Monitor, *Wifi location monitor accuracy*, Accessed: 2020-05-10, 2016.

[36] M. H. S. Eldaw, M. Levene, and G. Roussos, "Presence analytics: Making sense of human social presence within a learning environment," in *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*, IEEE, 2018, pp. 174–183.

[37] S. Ware, C. Yue, R. Morillo, J. Lu, C. Shang, J. Kamath, A. Bamis, J. Bi, A. Russell, and B. Wang, "Large-scale automatic depression screening using meta-data from wifi infrastructure," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, pp. 1–27, 2018.

[38] V. Das Swain, H. Kwon, B. Saket, M. B. Morshed, K. Tran, D. Patel, Y. Tian, J. Philipose, Y. Cui, T. Plötz, *et al.*, "Leveraging wifi network logs to infer social interactions: A case study of academic performance and student behavior," *arXiv e-prints*, 2020.

[39] E. S. Gurley, *Strategies to support the covid-19 response in lmics*, Accessed: 2021-02-03, 2020.

[40] G. I. of Technology, *Georgia tech launches campus coronavirus testing*, Available at: https://health.gatech.edu/coronavirus/testing-launched, 2020.

[41] U. of Illinois at Urbana-Champaign, *On-campus covid-19 testing*, Available at: https://covid19.illinois.edu/on-campus-covid-19-testing-data-dashboard/, 2020.

[42] B. University of California, *Coronavirus dashboard testing*, Available at: https://coronavirus.berkeley.edu/dashboard/, 2020.

[43] G. D. of Public Health, *Georgia department of public health daily status report*, Available at: https://dph.georgia.gov/covid-19-daily-status-report, 2020.

[44] C.-U. P. H. Distric, *Champaign-urbana covid-19 coronavirus information*, Available at: https://www.c-uphd.org/champaign-urbana-illinois-coronavirus-information.html, 2020.

[45] A. C. P. H. Department, *Real-time data of the impact of covid-19*, Available at: https://covid-19.acgov.org/data.page, 2020.

[46] S. Venkatramanan, A. Sadilek, A. Fadikar, C. L. Barrett, M. Biggerstaff, J. Chen, X. Dotiwalla, P. Eastham, B. Gipson, D. Higdon, *et al.*, "Forecasting influenza activity using machine-learned mobility map," *Nature Communications*, vol. 12, no. 1, pp. 1–12, 2021.

[47] S. M. Moghadas, M. C. Fitzpatrick, P. Sah, A. Pandey, A. Shoukat, B. H. Singer, and A. P. Galvani, "The implications of silent transmission for the control of covid-19 outbreaks," *Proceedings of the National Academy of Sciences*, vol. 117, no. 30, pp. 17 513–17 515, 2020.

[48] J. H. Stone, M. J. Frigault, N. J. Serling-Boyd, A. D. Fernandes, L. Harvey, A. S. Foulkes, N. K. Horick, B. C. Healy, R. Shah, A. M. Bensaci, *et al.*, "Efficacy of tocilizumab in patients hospitalized with covid-19," *New England Journal of Medicine*, vol. 383, no. 24, pp. 2333–2344, 2020.

[49] G. Gibson, J. S. Weitz, M. P. Shannon, B. Holton, A. Bryksin, B. Liu, S. Bramblett, J. Williamson, M. Farrell, A. Ortiz, C. T. Abdallah, and A. J. Garcıa, "Surveillance-to-diagnostic testing program for asymptomatic sars-cov-2 infections on a large, urban campus - georgia institute of technology, fall 2020," *medRxiv*, 2021, Available at: https://www.medrxiv.org/content/early/2021/01/31/2021.01.28.21250700. eprint: https://www.medrxiv.org/content/early/2021/01/31/2021.01.28.21250700.full.pdf.

[50] K. I. McKinnon, "Convergence of the nelder–mead simplex method to a nonstationary point," *SIAM Journal on optimization*, vol. 9, no. 1, pp. 148–158, 1998.

[51] K. Gaythorpe, N. Imai, G. Cuomo-Dannenburg, M. Baguelin, S. Bhatia, A. Boonyasiri, and A. Cori, *Report 8: Symptom progression of covid-19*, 2020.

[52] C. for Disease Control and Prevention, *Community npis: Flu prevention in community settings*, Available at: https://www.cdc.gov/nonpharmaceutical-interventions/community/index.html, 2020.

[53] A. Nierenberg and A. Pasick, *Schools briefing: University outbreaks and parental angst*, Available at: https://www.nytimes.com/2020/08/19/us/colleges-closing-covid.html, 2020.

[54] M. Andersen, "Early evidence on social distancing in response to covid-19 in the united states," *Available at SSRN 3569368*, 2020.

[55] H. S. Badr, H. Du, M. Marshall, E. Dong, M. M. Squire, and L. M. Gardner, "Association between mobility patterns and covid-19 transmission in the usa: A mathematical modelling study," *The Lancet Infectious Diseases*, vol. 20, no. 11, pp. 1247–1254, 2020.

[56] A. Smalley, "Higher education responses to coronavirus (covid-19)," in *National Conference of State Legislatures.[Accessed May 15, 2020]. https://www. ncsl. org/research/educatio education-responses-to-coronavirus-covid-19. aspx*, 2020.

[57] J. P. Azevedo, A. Hasan, D. Goldemberg, S. A. Iqbal, and K. Geven, *Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates*. The World Bank, 2020.

[58] E. Dorn, B. Hancock, J. Sarakatsannis, and E. Viruleg, "Covid-19 and student learning in the united states: The hurt could last a lifetime," *McKinsey & Company*, 2020.

[59] I. Chirikov, K. M. Soria, B. Horgos, and D. Jones-White, "Undergraduate and graduate students' mental health during the covid-19 pandemic," 2020.

[60] C. Woolston, *Signs of depression and anxiety soar among us graduate students during pandemic*, Available at: https://www.nature.com/articles/d41586-020-02439-6, 2020.

[61] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web.," Stanford InfoLab, Tech. Rep., 1999.

[62] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[63] P. Bonacich, "Some unique properties of eigenvector centrality," *Social networks*, vol. 29, no. 4, pp. 555–564, 2007.

[64] M. E. Newman, "Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality," *Physical review E*, vol. 64, no. 1, p. 016 132, 2001.

[65] L. Y. Saltzman, T. C. Hansel, and P. S. Bordnick, "Loneliness, isolation, and social support factors in post-covid-19 mental health.," *Psychological Trauma: Theory, Research, Practice, and Policy*, 2020.

[66] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[67] D. Dave, A. Friedson, K. Matsuzawa, J. J. Sabia, and S. Safford, "Jue insight: Were urban cowboys enough to control covid-19? local shelter-in-place orders and coronavirus case growth," *Journal of urban economics*, p. 103 294, 2020.

[68] M. E. Loades, E. Chatburn, N. Higson-Sweeney, S. Reynolds, R. Shafran, A. Brigden, C. Linney, M. N. McManus, C. Borwick, and E. Crawley, "Rapid systematic review: The impact of social isolation and loneliness on the mental health of children and adolescents in the context of covid-19," *Journal of the American Academy of Child & Adolescent Psychiatry*, 2020.

[69] B. Pfefferbaum and C. S. North, "Mental health and the covid-19 pandemic," *New England Journal of Medicine*, vol. 383, no. 6, pp. 510–512, 2020.

[70] D. Shah and T. Zaman, "Rumor centrality: A universal source detector," in *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of Computer Systems*, 2012, pp. 199–210.

[71] T. Lappas, E. Terzi, D. Gunopulos, and H. Mannila, "Finding effectors in social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2010, pp. 1059–1068.

[72] B. A. Prakash, J. Vreeken, and C. Faloutsos, "Efficiently spotting the starting points of an epidemic in a large graph," *Knowledge and information systems*, vol. 38, no. 1, pp. 35–59, 2014.

[73] B. Adhikari, Y. Zhang, S. E. Amiri, A. Bharadwaj, and B. A. Prakash, "Propagation-based temporal network summarization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 4, pp. 729–742, 2017.

[74] S. Sundareisan, J. Vreeken, and B. A. Prakash, "Hidden hazards: Finding missing nodes in large graph epidemics," in *Proceedings of the 2015 SIAM International Conference on Data Mining*, SIAM, 2015, pp. 415–423.

[75] P. Rozenshtein, A. Gionis, B. A. Prakash, and J. Vreeken, "Reconstructing an epidemic over time," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1835–1844.

[76] M. Charikar, C. Chekuri, T.-Y. Cheung, Z. Dai, A. Goel, S. Guha, and M. Li, "Approximation algorithms for directed steiner problems," *Journal of Algorithms*, vol. 33, no. 1, pp. 73–91, 1999.

[77] H. Shao, K. Hossain, H. Wu, M. Khan, A. Vullikanti, B. A. Prakash, M. Marathe, and N. Ramakrishnan, "Forecasting the flu: Designing social network sensors for epidemics," *arXiv preprint arXiv:1602.06866*, 2016.

[78] Y. Zhang and B. A. Prakash, "Data-aware vaccine allocation over large networks," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 10, no. 2, pp. 1–32, 2015.

[79] A. H. Cheetham and J. E. Hazel, "Binary (presence-absence) similarity coefficients," *Journal of Paleontology*, pp. 1130–1136, 1969.

[80] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, *et al.*, "Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand,"

[81] S. S. Morse, R. L. Garwin, and P. J. Olsiewski, "Next flu pandemic: What to do until the vaccine arrives?," 2006.

[82] N. Haug, L. Geyrhofer, A. Londei, E. Dervic, A. Desvars-Larrive, V. Loreto, B. Pinior, S. Thurner, and P. Klimek, "Ranking the effectiveness of worldwide covid-19 government interventions," *Nature human behaviour*, vol. 4, no. 12, pp. 1303–1312, 2020.

[83] Y. Zhang, B. Li, and K. Ramayya, "Learning individual behavior using sensor data: The case of gps traces and taxi drivers," *Forthcoming in Information Systems Research*, 2020.

[84] J. L. Wang and M. C. Loui, "Privacy and ethical issues in location-based tracking systems," in *2009 IEEE International Symposium on Technology and Society*, IEEE, 2009, pp. 1–4.

[85] E. Bagdasaryan, G. Berlstein, J. Waterman, E. Birrell, N. Foster, F. B. Schneider, and D. Estrin, "Ancile: Enhancing privacy for ubiquitous computing with use-based privacy," in *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society*, 2019, pp. 111–124.

[86] A. Pfitzmann and M. Hansen, *A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management*, 2010.

[87] G. A. Bruneau, V. C. Müller, and M. S. Gilthorpe, "The ethical imperatives of the covid 19 pandemic: A review from data ethics," *Veritas: Revista de Filosofía y Teología*, vol. 46, pp. 13–35, 2020.

[88] M. D. Fox, D. C. Bailey, M. D. Seamon, and M. L. Miranda, "Response to a covid-19 outbreak on a university campus—indiana, august 2020," *Morbidity and Mortality Weekly Report*, vol. 70, no. 4, p. 118, 2021.

[89] L. J. Schoen, "Guidance for building operations during the covid-19 pandemic," *ASHRAE Journal*, vol. 5, no. 3, 2020.

[90] S. Saul and S. Hubler, *Colleges vowed a safer spring. then students, and variants, arrived*, Available at: https://www.nytimes.com/2021/02/09/us/colleges-covid.html, 2021.

[91] M. Korn, *Colleges begin mapping out a more normal fall—with caveats*, Available at: https://www.wsj.com/articles/colleges-begin-mapping-out-a-more-normal-fallwith-caveats-11615800600?reflink=desktopwebshare_permalink, 2021.

[92] R. C. Rabin, *C.d.c. officials say most available evidence indicates schools can be safe if precautions are taken on campus and in the community*, Available at: https://www.nytimes.com/2021/01/26/world/cdc-schools-reopening.html, 2021.

[93] H. Zhao and Z. Feng, "Staggered release policies for covid-19 control: Costs and benefits of relaxing restrictions by age and risk," *Mathematical biosciences*, vol. 326, p. 108 405, 2020.

[94] J. C. Benneyan, C. Gehrke, I. Ilies, and N. Nehls, "Potential community and campus covid-19 outcomes under university and college reopening scenarios," *medRxiv*, 2020.

[95] S. G. Benzell, A. Collis, and C. Nicolaides, "Rationing social contact during the covid-19 pandemic: Transmission risk and social benefits of us locations," *Proceedings of the National Academy of Sciences*, vol. 117, no. 26, pp. 14 642–14 644, 2020.

[96] J. Suh, E. Horvitz, R. W. White, and T. Althoff, "Population-scale study of human needs during the covid-19 pandemic: Analysis and implications," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, pp. 4–12.

[97] S. J. Mooney and V. Pejaver, "Big data in public health: Terminology, machine learning, and privacy," *Annual review of public health*, vol. 39, pp. 95–112, 2018.